Lecture Notes in Networks and Systems 854

Pandian Vasant ·
Mohammad Shamsul Arefin ·
Vladimir Panchenko · J. Joshua Thomas ·
Elias Munapo · Gerhard-Wilhelm Weber ·
Roman Rodriguez-Aguilar *Editors*

Intelligent Computing and Optimization

Proceedings of the 6th International Conference on Intelligent Computing and Optimization 2023 (ICO2023), Volume 4

Springer

Lecture Notes in Networks and Systems 854

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Pandian Vasant · Mohammad Shamsul Arefin ·
Vladimir Panchenko · J. Joshua Thomas ·
Elias Munapo · Gerhard-Wilhelm Weber ·
Roman Rodriguez-Aguilar
Editors

Intelligent Computing and Optimization

Proceedings of the 6th International Conference on Intelligent Computing and Optimization 2023 (ICO2023), Volume 4

Springer

Editors
Pandian Vasant
Faculty of Electrical and Electronics Engineering, Modeling Evolutionary Algorithms Simulation and Artificial Intelligence
Ton Duc Thang University
Ho Chi Minh City, Vietnam

Vladimir Panchenko
Laboratory of Non-traditional Energy Systems, Department of Theoretical and Applied Mechanics, Federal Scientific Agroengineering Center VIM
Russian University of Transport
Moscow, Russia

Elias Munapo
School of Economics and Decision Sciences
North West University
Mmabatho, South Africa

Roman Rodriguez-Aguilar
Facultad de Ciencias Económicas y Empresariales, School of Economic and Business Sciences
Universidad Panamericana
Mexico City, Mexico

Mohammad Shamsul Arefin
Department of Computer Science
Chittagong University of Engineering and Technology
Chittagong, Bangladesh

J. Joshua Thomas
Department of Computer Science
UOW Malaysia KDU Penang University College
George Town, Malaysia

Gerhard-Wilhelm Weber
Faculty of Engineering Management
Poznań University of Technology
Poznan, Poland

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-50150-0 ISBN 978-3-031-50151-7 (eBook)
https://doi.org/10.1007/978-3-031-50151-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

The sixth edition of the *International Conference on Intelligent Computing and Optimization (ICO'2023) was held during April 27–28, 2023, at G Hua Hin Resort and Mall, Hua Hin, Thailand*. The objective of the international conference is to bring the global research scholars, experts and scientists in the research areas of intelligent computing and optimization from all over the world to share their knowledge and experiences on the current research achievements in these fields. This conference provides a golden opportunity for global research community to interact and share their novel research results, findings and innovative discoveries among their colleagues and friends. The proceedings of ICO'2023 is published by SPRINGER (in the book series ***Lecture Notes in Networks and Systems***) and indexed by **SCOPUS**.

Almost 70 authors submitted their full papers for the 6th ICO'2023. They represent more than 30 countries, such as Australia, Bangladesh, Bhutan, Botswana, Brazil, Canada, China, Germany, Ghana, Hong Kong, India, Indonesia, Japan, Malaysia, Mauritius, Mexico, Nepal, the Philippines, Russia, Saudi Arabia, South Africa, Sri Lanka, Thailand, Turkey, Ukraine, UK, USA, Vietnam, Zimbabwe and others. This worldwide representation clearly demonstrates the growing interest of the global research community in our conference series. The organizing committee would like to sincerely thank all the authors and the reviewers for their wonderful contribution for this conference. The best and high-quality papers will be selected and reviewed by International Program Committee in order to publish the extended version of the paper in the international indexed journals by SCOPUS and ISI WoS.

This conference could not have been organized without the strong support and help from LNNS SPRINGER NATURE, Easy Chair, IFORS and the Committee of ICO'2023. We would like to sincerely thank *Prof. Roman Rodriguez-Aguiler* (Universidad Panamericana, Mexico) and *Prof. Mohammad Shamsul Arefin* (Daffodil International University, Bangladesh), *Prof. Elias Munapo* (North West University, South Africa) and *Prof. José Antonio Marmolejo Saucedo* (National Autonomous University of Mexico, Mexico) for their great help and support for this conference.

We also appreciate the wonderful guidance and support from *Dr. Sinan Melih Nigdeli* (Istanbul University—Cerrahpaşa, Turkey), *Dr. Marife Rosales* (Polytechnic University of the Philippines, Philippines), *Prof. Rustem Popa* (Dunarea de Jos University, Romania), *Prof. Igor Litvinchev* (Nuevo Leon State University, Mexico), *Dr. Alexander Setiawan* (Petra Christian University, Indonesia), *Dr. Kreangkri Ratchagit* (Maejo University, Thailand), *Dr. Ravindra Boojhawon* (University of Mauritius, Mauritius), *Prof. Mohammed Moshiul Hoque* (CUET, Bangladesh), *Er. Aditya Singh* (Lovely Professional University, India), *Dr. Dmitry Budnikov* (Federal Scientific Agroengineering Center VIM, Russia), *Dr. Deepanjal Shrestha* (Pokhara University, Nepal), *Dr. Nguyen Tan Cam* (University of Information Technology, Vietnam) and *Dr. Thanh Dang Trung* (Thu Dau Mot University, Vietnam). The ICO'2023 committee would like to sincerely thank all the authors, reviewers, **keynote speakers** (*Prof. Roman Rodriguez-Aguiler,* *Prof. Kaushik Deb, Prof. Rolly Intan, Prof. Francis Miranda, Dr. Deepanjal Shrestha, Prof. Sunarin Chanta*), **plenary speakers** (*Prof. Celso C. Ribeiro, Prof. José Antonio Marmolejo, Dr. Tien Anh Tran*), session chairs and participants for their outstanding contribution to the success of the 6th ICO'2023 in Hua Hin, Thailand.

Finally, we would like to sincerely thank *Prof. Dr. Janusz Kacprzyk, Dr. Thomas Ditzinger, Dr. Holger Schaepe* and *Ms. Varsha Prabakaran* of **LNNS SPRINGER NATURE** for their great support, motivation and encouragement in making this event successful in the global stage.

April 2023

Dr. Pandian Vasant (Chair)
Prof. Dr. Gerhard-Wilhelm Weber
Prof. Dr. Mohammad Shamsul Arefin
Prof. Dr. Roman Rodriguez-Aguiler
Dr. Vladimir Panchenko
Prof. Dr. Elias Munapo
Dr. J. Joshua Thomas

Contents

About the Editors

Pandian Vasant is Research Associate at Modeling Evolutionary Algorithms Simulation and Artificial Intelligence, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam, and Editor in Chief of *International Journal of Energy Optimization and Engineering* (IJEOE). He holds Ph.D. in Computational Intelligence (UNEM, Costa Rica), M.Sc. (University Malaysia Sabah, Malaysia, Engineering Mathematics) and B.Sc. (Hons, Second Class Upper) in Mathematics (University of Malaya, Malaysia). His research interests include soft computing, hybrid optimization, innovative computing and applications. He has co-authored research articles in journals, conference proceedings, presentations, special issues Guest Editor, chapters (312 publications indexed in Research-Gate) and General Chair of EAI International Conference on Computer Science and Engineering in Penang, Malaysia (2016) and Bangkok, Thailand (2018). In the years 2009 and 2015, he was awarded top reviewer and outstanding reviewer for the journal Applied Soft Computing (Elsevier). He has 35 years of working experience at the universities. Currently, Dr. Pandian Vasant is General Chair of the International Conference on Intelligent Computing and Optimization (https://www.icico.info/) and Research Associate at Modeling Evolutionary Algorithms Simulation and Artificial Intelligence, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, HCMC, Vietnam.

Professor Dr. Mohammad Shamsul Arefin is in lien from Chittagong University of Engineering and Technology (CUET), Bangladesh and currently affiliated with the Department of Computer Science and Engineering (CSE), Daffodil International University (DIU), Dhaka, Bangladesh. Earlier he was the head of CSE Department, CUET. Prof. Arefin received his Doctor of Engineering Degree in Information Engineering from Hiroshima University, Japan with support of the scholarship of MEXT, Japan. As a part of his doctoral research, Dr. Arefin was with IBM Yamato Software Laboratory, Japan. His research includes data privacy and mining, big data management, IoT, Cloud Computing, Natural Language processing, Image Information Processing, Social Networks Analysis and Recommendation Systems and IT for agriculture, education and environment. Prof. Arefin is the Editor in Chief of *Computer Science and Engineering Research Journal* (ISSN: 1990-4010) and was the Associate Editor of *BCS Journal of Computer and Information Technology* (ISSN: 2664-4592) and a reviewer as well as TPC member of many international journals and conferences. Dr. Arefin has more than 120 referred publications in international journals, book series and conference proceedings. He delivered more than 30 keynote speeches/invited talks. He also received a good number of research grants/funds from home and abroad. Dr. Arefin is a senior member of IEEE, Member of ACM, Fellow of IEB and BCS. Prof. Arefin involves/earlier involved in many professional activities such as Chairman of Bangladesh Computer Society (BCS) Chittagong Branch; Vice-President (Academic) of BCS National Committee; Executive Committee Member of IEB Computer Engineering Division; Advisor, Bangladesh Robotic Foundation. He was also a member of pre-feasibility study team of CUET IT Business Incubator, first campus based IT Business Incubator in Bangladesh. Prof. Arefin is an Principle Editor of the **Lecture Notes on Data Engineering and Communications Technologies book series** (LNDECT, Volume 95) published by Springer and an editor of the books on *Applied Informatics for Industry 4.0*, *Applied Intelligence for Industry 4.0 and Computer Vision and Image Analysis for Industry 4.0* to be published Tailor and Francis. Prof. Arefin is the Vice-Chair (Technical) of IEEE CS BDC for the year 2022. He was the Vice-Chair (Activity) of IEEE CS BDC for the year 2021 and the Conference Co-Coordinator of IEEE CS BDC for two consecutive years, 2018 and 2019. He is acting as a TPC Chair of MIET 2022 and the TPC Chair of IEEE Summer Symposium 2022. He was the Organizing Chair of International Conference on Big Data, IoT and Machine Learning (BIM 2021) and National Workshop on Big Data and Machine Learning (BDML 2020). He served as the TPC Chair, International Conference on Electrical, Computer and Communication Engineering (ECCE 2017); Organizing Co-chair, ECCE 2019, Technical Co-chair, IEEE CS BDC Winter Symposium 2020 and Technical Secretary, IEEE CS BDC Winter Symposium 2021. Dr. Arefin helped different international conferences in the form of track chair, TPC member, reviewer and/or secession chair etc. He is a reviewer of many reputed journals including *IEEE Access, Computing Informatics, ICT Express, Cognitive Computation* etc. Dr. Arefin visited Japan, Indonesia, Malaysia, Bhutan, Singapore, South Korea, Egypt, India, Saudi Arabia and China for different professional and social activities.

Vladimir Panchenko is an Associate Professor of the "Department of Theoretical and Applied Mechanics" of the "Russian University of Transport", Senior Researcher of the "Laboratory of Non-traditional Energy Systems" of the "Federal Scientific Agroengineering Center VIM" and the Teacher of additional education. Graduated from the "Bauman Moscow State Technical University" in 2009 with the qualification of an engineer. Ph.D. thesis of the specialty "Power plants based on renewable energy" was defended in 2013. From 2014 to 2016 Chairman of the Council of Young Scientists and the Member of the Academic Council of the All-Russian Institute for Electrification of Agriculture, Member of the Council of Young Scientists of the Russian University of Transport, Member of the International Solar Energy Society, Individual supporter of Greenpeace and the World Wildlife Fund, Member of the Russian Geographical Society, Member of the Youth section of the Council "Science and Innovations of the Caspian Sea", Member of the Committee on the use of renewable energy sources of the Russian Union of Scientific and Engineering Public Associations. Diplomas of the winner of the competition of works of young scientists of the All-Russian Scientific Youth School with international participation "Renewable Energy Sources", Moscow State University M.V. Lomonosov in 2012, 2014, 2018 and 2020, Diploma with a bronze medal of the 15th Russian agro-industrial exhibition "Golden Autumn—2013", Diploma with a gold medal of the 18th Russian agro-industrial exhibition "Golden Autumn—2016", Diploma with a silver medal of the XIX Moscow International Salon of Inventions and Innovative technologies "Archimedes—2016", Diploma for the winning the schoolchildren who have achieved high results in significant events of the Department of Education and Science of the City of Moscow (2020–2021, School No. 2045). Scientific adviser of schoolchildren-winners and prize-winners of the Project and Research Competition "Engineers of the Future" at NUST MISiS 2021 and

RTU MIREA 2022. Invited expert of the projects of the final stages of the "Engineers of the Future" (2021, 2022) and the projects of the "Transport of the Future" (2022, Russian University of Transport). Grant "Young teacher of MIIT" after competitive selection in accordance with the Regulations on grants for young teachers of MIIT (2016–2019). Scholarship of the President of the Russian Federation for 2018–2020 for young scientists and graduate students carrying out promising research and development in priority areas of modernization of the Russian economy. Grant of the Russian Science Foundation 2021 "Conducting fundamental scientific research and exploratory scientific research by international research teams". Reviewer of articles, chapters and books IGI, Elsevier, Institute of Physics Publishing, *International Journal of Energy Optimization and Engineering, Advances in Intelligent Systems and Computing, Journal of the Operations Research Society of China, Applied Sciences, Energies, Sustainability, AgriEngineering, Ain Shams Engineering Journal, Concurrency and Computation: Practice and Experience.* Presenter of the sections of the Innovations in Agriculture conference, keynote speaker of the ICO 2019 conference session, keyspeaker of the special session of the ICO 2020 conference. Assistant Editor since 2019 of the "International Journal of Energy Optimization and Engineering", Guest Editor since 2019 of the Special Issues of the journal MDPI (Switzerland) "Applied Sciences", Editor of the book of the "IGI GLOBAL" (USA), as well as book of the "Nova Science Publisher" (USA). Participated in more than 100 exhibitions and conferences of various levels. Published more than 250 scientific papers, including 14 patents, 1 international patent, 6 educational publications, 4 monographs.

J. Joshua Thomas is an Associate Professor at UOW Malaysia KDU Penang University College, Malaysia since 2008. He obtained his Ph.D. (Intelligent Systems Techniques) in 2015 from University Sains Malaysia, Penang, and Master's degree in 1999 from Madurai Kamaraj University, India. From July to September 2005, he worked as a research assistant at the Artificial Intelligence Lab in University Sains Malaysia. From March 2008 to March 2010, he worked as a research associate at the same University. Currently, he is working with Machine Learning, Big Data, Data Analytics, Deep Learning, specially targeting on

Convolutional Neural Networks (CNN) and Bi-directional Recurrent Neural Networks (RNN) for image tagging with embedded natural language processing, End to end steering learning systems and GAN. His work involves experimental research with software prototypes and mathematical modelling and design He is an editorial board member for the *Journal of Energy Optimization and Engineering* (IJEOE), and invited guest editor for *Journal of Visual Languages Communication* (JVLC-Elsevier). Recently with *Computer Methods and Programs in Biomedicine* (Elsevier). He has published more than 40 papers in leading international conference proceedings and peer reviewed journals.

Elias Munapo has a Ph.D. obtained in 2010 from the National University of Science and Technology (Zimbabwe) and is a Professor of Operations Research at the North West University, Mafikeng Campus in South Africa. He is a Guest Editor of the *Applied Sciences Journal* and has co-published two books. The first book is titled *Some Innovations in OR Methodology: Linear Optimization* and was published by Lambert Academic publishers in 2018. The second book is titled *Linear Integer Programming: Theory, Applications, and Recent Developments* and was published by De Gruyter publishers in 2021. Professor Munapo has co-edited a number of books, is currently a reviewer of a number of journals, and has published over 100 journal articles and book chapters. In addition, Prof. Munapo is a recipient of the North West University Institutional Research Excellence award and is a member of the Operations Research Society of South Africa (ORSSA), EURO, and IFORS. He has presented at both local and international conferences and has supervised more than 10 doctoral students to completion. His research interests are in the broad area of operations research.

Gerhard-Wilhelm Weber is a Professor at Poznan University of Technology, Poznan, Poland, at Faculty of Engineering Management. His research is on mathematics, statistics, operational research, data science, machine learning, finance, economics, optimization, optimal control, management, neuro-, bio- and earth-sciences, medicine, logistics, development, cosmology and generalized space-time research. He is involved in the organization of scientific life internationally. He received Diploma and Doctorate in Mathematics, and Economics/Business Administration, at RWTH Aachen, and Habilitation at TU Darmstadt (Germany). He replaced Professorships at University of Cologne, and TU Chemnitz, Germany. At Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey, he was a Professor in Financial Mathematics and Scientific Computing, and Assistant to the Director, and has been a member of five further graduate schools, institutes and departments of METU. G.-W. Weber has affiliations at Universities of Siegen (Germany), Federation University (Ballarat, Australia), University of Aveiro (Portugal), University of North Sumatra (Medan, Indonesia), Malaysia University of Technology, Chinese University of Hong Kong, KTO Karatay University (Konya, Turkey), Vidyasagar University (Midnapore, India), Mazandaran University of Science and Technology (Babol, Iran), Istinye University (Istanbul, Turkey), Georgian International Academy of Sciences, at EURO (Association of European OR Societies) where he is "Advisor to EURO Conferences" and IFORS (International Federation of OR Societies), where he is member in many national OR societies, honorary chair of some EURO working groups, subeditor of IFORS Newsletter, member of IFORS Developing Countries Committee, of Pacific Optimization Research Activity Group, etc. G.-W. Weber has supervised many M.Sc. and Ph.D. students, authored and edited numerous books and articles, and given many presentations from a diversity of areas, in theory, methods and practice. He has been a member of many international editorial, special issue and award boards; he participated at numerous research projects; he received various recognitions by students, universities, conferences and scientific organizations. G.-W. Weber is an IFORS Fellow.

Roman Rodriguez-Aguilar is a professor in the School of Economic and Business Sciences of the "Universidad Panamericana" in Mexico. His research is on large-scale mathematical optimization, evolutionary computation, data science, statistical modeling, health economics, energy, competition, and market regulation. He is particularly interested in topics related to artificial intelligence, digital transformation, and Industry 4.0. He received his Ph.D. at the School of Economics at the National Polytechnic Institute, Mexico. He also has a master's degree in Engineering from the School of Engineering at the National University of Mexico (UNAM), a master's degree in Administration and Public Policy in the School of Government and Public Policy at Monterrey Institute of Technology and Higher Education, a postgraduate in applied statistics at the Research Institute in Applied Mathematics and Systems of the UNAM and his degree in Economics at the UNAM. Prior to joining Panamericana University, he has worked as a specialist in economics, statistics, simulation, finance, and optimization, occupying different management positions in various public entities such as the Ministry of Energy, Ministry of Finance, and Ministry of Health. At present, he has the second-highest country-wide distinction granted by the Mexican National System of Research Scientists for scientific merit (SNI Fellow, Level 2). He has co-authored research articles in science citation index journals, conference proceedings, presentations, and book chapters.

Fuzzy Logic, ANN, Green Cloud Computing, and Smart Algorithms

Time Series Analysis in COVID-19 Daily Reported Cases in South Africa: A Box-Jenkins Methodology

Happy Motene[1], Phemelo Seaketso[1], Elias Munapo[1](✉), and Precious Mdlongwa[2]

[1] Department of Business Statistics and Operations Research, School of Economic Sciences, North West University, Mafikeng Campus, Mahikeng, South Africa
elias.munapo@nwu.ac.za

[2] National University of Science & Technology, PO Box AC 939, Ascot, Bulawayo, Zimbabwe

Abstract. COVID-19 has affected the lives of South Africans and the world at large. In this study, daily COVID-19-reported cases are considered. The COVID-19 pandemic has affected healthcare facilities and all other economic structures in South Africa. Forecasting daily COVID-19 Cases is helpful and reduces the pressure of uncertainty this pandemic has and the understanding of its nature. In this study, ARIMA and SARIMA models are used to understand the nature and patterns of the COVID-19 pandemic The R Software is used in the analysis and the models are selected using the AIC and BIC criteria by observing the ACF and PACF of each of the models. This study proposes that forecasting can help the South African government and healthcare officials to understand the nature of COVID-19 and observe its effects on the economy. The Box-Jenkins methodology is applied using R Software so to choose the right model and forecast for the next 60 days. We hope decisions that will be made will help manage and control the COVID-19 pandemic.

Keywords: COVID-19 · Forecasting · ARIMA · SARIMA · Box and Jenkins

Acronyms

The following list of acronyms are used

AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
COVID-19	Coronavirus disease 2019
AR	Autoregressive
MA	Moving Average
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
D/d	Differencing
ACF	Autocorrelated Function

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 3–11, 2024.
https://doi.org/10.1007/978-3-031-50151-7_1

PACF Partial Autocorrelated Function
NAAT Nucleic Acid Amplification Test
NNAR Neural network Nonlinear AutoRegressive (NNAR)

1 Introduction

1.1 Background to Study

The 2019 COVID-19 pandemic has caused widespread devastation. COVID-19 can respiratorily infect human cells (Ciotti et al. 2020). In less than two years, the virus has killed 4.55 million people worldwide, including 85 952 in South Africa (Worldometer 2021). In spite of the high number of fatalities, South Africa reported an average of 4,020 cases of COVID-19 per week by August 2021.

According to Worldometer (2021), as of September 18, 2021, South Africa had 2.88 million active cases, which is a large number that could lead to a shortage of medical facilities. Since the COVID-19 pandemic began, South Africa has encouraged people to wear face masks. Keeping the mouth and nose covered prevents virus transmission (WHO 2020). South Africa implemented five levels of lockdown to combat the virus. The Box and Jenkins concept is used to analyse COVID-19 data in this study (1970). The time series approach developed by Box and Jenkins, which forecasts univariate time series data, is the most widely used (Box et al. 2016).

1.2 Literature Review

Articles on COVID-19 predictability are reviewed in this section. In the same section, the ARIMA and SARIMA models are explained. Examining the origins of COVID-19 can reveal pandemic factors. Understanding the pandemic aids in the identification of random factors influencing its spread.

COVID-19's devastation continues. South Africa waited until March 18, 2020, when the first 54 COVID-19 cases were officially reported. Wuhan, China, discovered COVID-19. It infects human cells (Ciotti et al. 2020). The World Health Organization (2020) recommends looking for three signs to confirm a COVID-19 diagnosis. Case 1 is suspected, Case 2 is probable, and Case 3 is confirmed by a positive NAAT test at the World Health Organization (2020).

The process of predicting future events is known as forecasting (Montgomery et al. 2015). The Autoregressive Integrated Moving Average (ARIMA) is a predictive model that is frequently utilised for the purpose of forecasting time series data (Demir and Kirisci 2021). Based on data observations, COVID-19 has a pattern of recurring or seasonal behaviour. Unfortunately, time series with seasonal components do not respond well to the ARIMA model. The pandemic, on the other hand, is able to be investigated with the help of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The SARIMA concept is one that displays or identifies the seasonal components that are present in data.

In contrast to SARIMA, which produced unsatisfactory model results, ARIMA had a better forecast model of the number of infected people, according to Abolmaali and

Shirzaei (2021). Despite the fact that ARIMA can handle almost any time series data set, it suffers when dealing with seasonal data. However, SARIMA has drawbacks and has not outperformed competing predictive models like NNAR (Demir and Kirisci 2021). Infection with COVID-19 may cause very few or no symptoms at all. There is a wide range of severity in both the symptoms and the signs. The incubation period for COVID-19 could last for two weeks or even longer (Gebretensae and Asmelash 2021). In light of the limited resources available to combat the pandemic, it will be helpful for the government and the health department to make plans based on a forecast of a COVID-19 pattern. This will be of assistance to health departments in their planning for hospital beds and ventilators, both of which will be required when providing care to a patient suffering from COVID-19. In addition, the government can determine when it is appropriate to implement lockdown regulations by being aware of the peak periods identified by the time series. To control the COVID-19 pandemic, it is crucial to comprehend the disease's potential for rapid spread. As a result, time series forecasting is a crucial component of this investigation, as it has been utilised effectively in the past.

Several disciplines, including economics, finance, engineering, and metrology, use time series analysis. Today, the most popular field of study is time series prediction, which includes electricity prediction (Hu and Chen 2020). Time series analyses were used in a Dynamic Weight Ensemble Model-based electricity consumption forecast system (DWEN). The DWEN method predicts adaptive weights based on each model's previous day's performance. Hu and Chen (2020) demonstrated that the method's estimated results were more accurate than the actual power consumption.

2 Research Methodology

The main goal of this study is to predict daily COVID-19 infection cases using time series data. The Box-Jenkins method and the ARIMA or SARIMA models will both be used to achieve this goal. The focus will be on COVID-19 infection cases that have been confirmed in order to assist the health department in understanding COVID-19 and developing a clear plan for treating COVID-19 infection cases that are confirmed on a daily basis.

The following are the study's sub-objectives, which are to:

1. determine the situation's nature in relation to the observations' chronological order,
2. investigate and identify the best model for projecting COVID-19 time series data, and
3. forecast the time series variable's future values.

To achieve these objectives, we formulate the following null hypotheses:

1. H_{01}: Covid-19 can be forecasted.
2. H_{02}: Covid-19 follows a seasonal pattern.
3. H_{03}: Residuals of the fitted model are white noise.

Even though South Africa has a large number of hospitals and other medical facilities that are of a high standard, the country could see an increase in the number of deaths that COVID-19 causes if the pandemic is not managed effectively. The COVID-19 outbreak serves as a sobering reminder of the ongoing issues brought on by re-emerging infectious pathogens and the requirement for vigilant research, prompt diagnosis, and constant monitoring in order to comprehend the fundamental biology of organisms, as well as our susceptibility to them, and to create efficient countermeasures (Demir and Kirisci 2021).The data for this study were analysed using R software, which is a statistical computation and graphics system (Baier and Neuwirth 2007). Figure 1 summarises the data analysis procedure for this study.

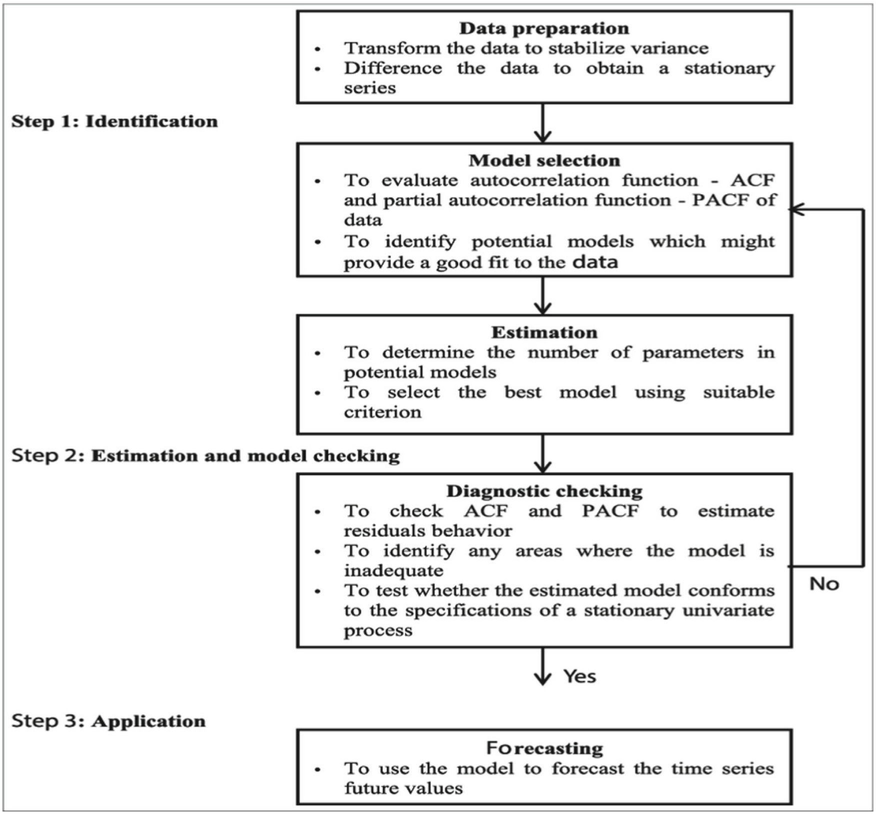


Fig. 1. A summary of data analysis steps.

The daily data used in this study came from Worldometer's website database (Worldometer 2021).

3 Data Analysis

3.1 Initial Stages of the Analysis

The selection of models makes use of the frequency of 7. This study uses the number of observations that were made before seasonal patterns to define the meaning of the term "frequency" (Hyndman and Athanasopoulos 2018). The original time plot as well as the first and second-time series plots are obtained. Autocorrelation and partial autocorrelation plots, the extended autocorrelation function table, auto-ARIMA, and model estimation follow. The fitted model is checked for adequacy and then used for forecasting. Finally, the analysis provides answers to hypotheses.

South Africa's daily confirmed COVID-19 cases are displayed in Fig. 2.

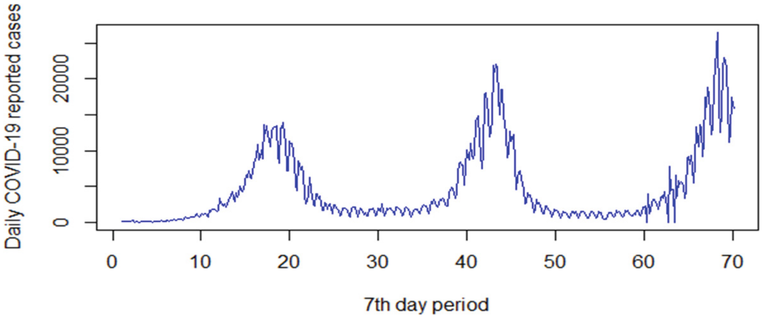


Fig. 2. Time series plot of data

Figure 2 shows a cyclical or seasonal pattern over the course of 485 days, or 70 weeks. As seen in Fig. 2, the COVID-19 reported cases plot fluctuates over the time period with an irregular increasing and decreasing trend. From 18 March 2020 to 16 July 2021, the graph reveals that South Africa experienced periodic or seasonal fluctuations. However, between Weeks 20 and 40 and again between Weeks 50 and 60, the country likely experienced one of the smallest declines. The series started to pick up steam from Week 60 to Week 70, and it will peak in 2021, probably in June or July. Winter in South Africa begins between Weeks 60 and 70, which led to an increase in COVID-19 reported cases.

The impact led to an increase in the number of fatalities and rendered the available space in hospitals inadequate for treating all of the patients. One of the African nations that were impacted negatively by these devastating effects was South Africa. On occasion, a slight decline is observed, such as around week 30 and weeks 50–60. This decline is not significant. This could be due to the nationwide lockdown that was implemented as a precaution against the further spread of COVID-19 throughout the country. Figure 2 also shows the seasonal variation.

In Fig. 3, the autocorrelations do not reach zero, indicating non-stationarity. The findings in Fig. 3 are consistent with the theoretical guidelines shown in Table 1. According to Box and Jenkins, this provides a solid foundation for concluding that the series is not stationary (1970). For this reason, the R *auto.arima* () function was used to obtain the suitable model. Some of the fitted models are shown in Table 1.

The two models that are tentatively considered are seasonal models, which is the category that both of these models belong. On the other hand, this suggests that the testing of hypotheses to see if the data follows a seasonal pattern is acceptable. We then choose the model with the lowest AIC and BIC values. Table 2 lists selected models.

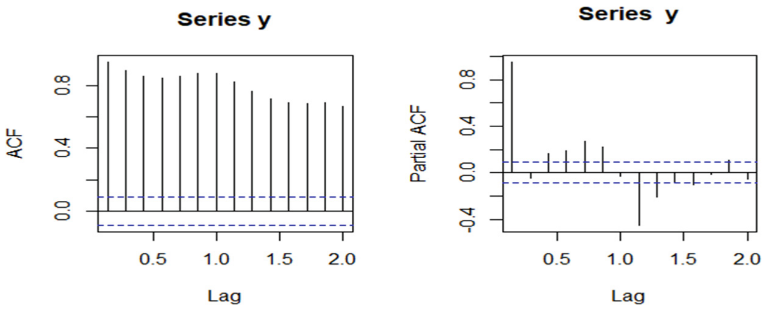


Fig. 3. ACF and PACF plots of the initial time series

Table 1. Some ARIMA models for South Africa's covid data

FIT	ARIMA Models	AIC	BIC
fit_sarma1	(4,1,1)(3,1,2)[7]	8139.08	8184.92
fit_sarma11	(1,1,2)(1,1,1)[7]	8140.96	8165.97
fit_sarma12	(3,1,2)(1,1,1)[7]	8144.73	8178.07
fit_sarma13	(4,1,0)(2,1,0)[7]	8164.95	8194.12
fit_sarma2	(1,0,2)(1,1,1)[7]	8151.00	8176.02
fit_sarma21	(1,1,2)(1,0,1)[7]	8257.80	8282.89
fit_sarma22	(3,1,2)(1,0,1)[7]	8261.58	8295.03
fit_sarma23	(4,1,0)(2,0,0)[7]	8253.58	8282.86

Table 2. Selected significant models

fit_sarma1	(4,1,1)(3,1,2)[7]	8139.08	8184.92
fit_sarma11	(1,1,2)(1,1,1)[7]	8140.96	8165.97

3.2 Diagnostic Checking

Figure 4 shows the histogram plots of the selected models residuals. The histogram plots in Fig. 4 confirm the normality within the data by displaying a symmetric shape.

The results of the Ljung Box test, which is used to determine whether or not the residuals of selected models exhibit autocorrelation, are shown in Table 3.

From Table 3, the null hypothesis for no significance autocorrelation is rejected for the sarma11 model whose p-value is less than 0.05 So, we choose (fit sarma1) as the model to use for forecasting, since its residuals have no autocorrelation.

3.3 Forecasting with ARIMA Model

Figure 5 shows forecasts for the next 60days.

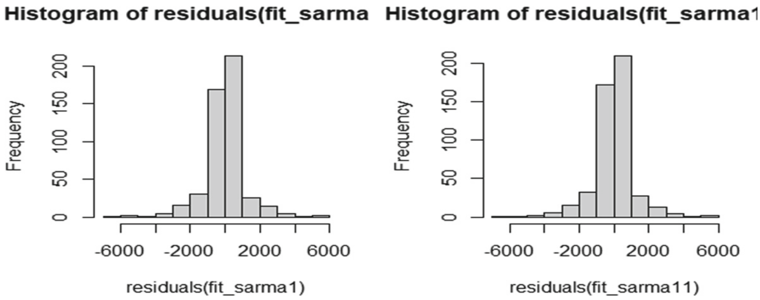


Fig. 4. Histogram plots of selected models

Table 3. Selected models Box-Ljung test

Data	X-squared	Df	p-value
resid(fit_sarma1	28.207	19	0.079520
resid(fit_sarma11)	38.688	19	0.004845

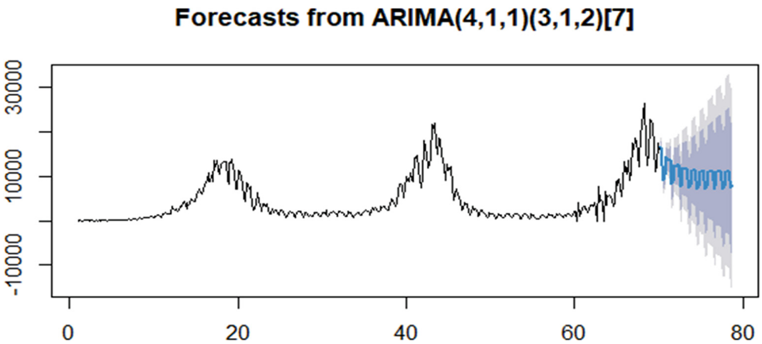


Fig. 5. SARIMA (4,1,1)(3,1,2)[7] forecast

According to Fig. 5, the number of confirmed cases is expected to go down seasonally during the 60-day period of the forecast (September 2021 to October 2021).

3.4 Hypothesis Testing Results

- We developed a model that we used to generate forecasts for covid 19. This supports our null hypothesis H_{01} that the covid 19 can be predicted.
- We fitted a seasonal model that supports the hypothesis that Covid-19 cases are seasonal i.e. H_{02}. The reported South African Covid-19 cases seem to follow a seasonal pattern, which is confirmed by the *auto.arima* () function in R, which shows that the data follow a weekly pattern. We conclude, therefore, that the data is seasonal.

- With a p-value of 0.079520 for the Box-Ljung test, there was no autocorrelation in the residuals of the model that was ultimately chosen. This is consistent with the hypothesis H_{03} that the selected residuals are white noise.

4 Conclusions, Recommendations & Limitations

4.1 Conclusion

COVID-19 cases are expected to decrease as a result of the vaccine's introduction, according to forecasting. Both models, (4,1,1)(3,1,2)[7] and (1,1,2)(1,1,1)[7], were significant, with the AIC and BIC having the lowest values of any model fitted. Based on well-behaved residuals with no autocorrelation, the best model was chosen as (4,1,1)(3,1,2) [7]. Major South African organizations could use the study's crucial evidence to decide how to stop the spread of COVID-19 and mitigate it.

4.2 Recommendations

The researchers advise the government to review the pandemic's fundamental behaviour. It is essential to have an accurate, current pandemic forecast. Weekly forecasting is required to understand potential external factors driving the spread of the virus. In this pandemic, short-term forecasts are more useful than long-term ones because there may be other factors at play that are increasing the number of reported cases we are unaware of.

4.3 Limitation

The study's limitation is that neither the demographics of reported cases nor the effects of the travel restrictions were examined.

References

Abolmaali, S., Shirzaei, S.: Forecasting COVID-19 number of cases by implementing ARIMA and SARIMA with grid search in the United States. Medrxiv (2021). https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Forecasting+Covid-19+number+of+cases+by+implementing+ARIMA+and+sarima+with+grid+search+in+United+States&btnG=#d=gs_qabs&u=%23p%3D998yCGGyt7QJ. Accessed 31 August 2021

Baier, T., Neuwirth, E.: Excel::COM. Comp. Statis. **22**(1), 91–108 (2007). https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Excel%3A%3ACom%3A%3AR&btnG=#d=gs_qabs&u=%23p%3D6NuLJ_bUtNMJ

Box, G., Reinsel, G., Jenkins, G., Ljung, G.: Time Series Analysis, 5th ed. Wisely¸ New Jersey, United States of America (2016), pp. 129–174

Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W., Wang, C., Bernardini, S.: The covid-19 pandemic. Tandfonline [online] **57**(6), 1–50 (2020). https://www.tandfonline.com/loi/ilab20. Accessed 31 August 2021

Demir, I., Kirisci, M.: Forecasting COVID-19 diseases cases using the sarima-nna hybrid model. Google Scholar **10.1101**(21256108), 1–10 (2021). https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Forecasting+Covid-19+diseases+cases+using+the+sarima-nna+hybrid+model&btnG=#d=gs_qabs&u=%23p%3DZG2yEosOA5EJ. Accessed 31 August 2021

Gebretensae, Y., Asmelash, D.: Trend analysis and forecasting the spread of COVID-19 pandemic in Ethiopia using Box–Jenkins modeling procedure. Int. J. General Med. [online] **14**, 1485–1498 (2021). https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Trend+Analysis+and+Forecasting+the+Spread+of+COVID-19+Pandemic+in+Ethiopia+Using+Box%E2%80%93Jenkins+Modeling+Procedure&btnG=#d=gs_qabs&u=%23p%3Dt9oCDWMPQzAJ

Hu, C., Chen, Y.: Forecasting time series for electricity consumption data using dynamic weight ensemble model. IEEE Xplore, 1–9 (2020)

Hyndman, R., Athanasopoulos, G.: Forecasting: Principles and Practice, 2nd ed. [online]. Otexts.com. https://otexts.com/fpp2/.Accessed 3 November 2021

Montgomery, D.C., Jennings, C.L., Kulahci, M.: Introduction to Time Series Analysis and Forecasting, 2nd edn. Wiley, New York (2015)

WHO: WHO COVID-19 case definition. Who.int. https://www.who.int/publications-detail-redirect/WHO-2019-nCoVSurveillance_Case_Definition-2020.2. Accessed 31 August 2021

Worldometer: 2021. https://www.worldometers.info/coronavirus/country/south-africa/. Accessed 31 August 2021

A Model of Continuous Investing in Information Security with Multifactory Accounting in a Fuzzy Statement

V. Malyukov[1], V. Lakhno[1], I. Malyukova[2], O. Kryvoruchko[3], A. Desiatko[3](✉), and M. Tsiutsiura[4]

[1] National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
lva964@nubip.edu.ua
[2] Rating Agency "Expert-Rating", Kyiv, Ukraine
[3] State University of Trade and Economics, Kyiv, Ukraine
ev_kryvorucko@ukr.net, desyatko@gmail.com
[4] Kyiv National University of Construction and Architecture, Kyiv, Ukraine

Abstract. The proposed model aims to manage the continuous investment process in information security systems (ISS) of information security objects (OIS) while considering multifactoriality in a fuzzy formulation. Unlike existing solutions, the model takes into account continuous investment in the protection of information security. A new class of bilinear differential games in a fuzzy formulation is described in the solution process. The novelty of the model lies in the fact that for bilinear differential games Pontryagin, namely the first direct method and the alternating integral method. Also, the methods of the school of Krasovsky, despite the fact that the condition for the existence of a saddle point for a "small" game is satisfied here. Therefore, a new solution method was proposed—discrete-approximation, which made it possible to solve the problem under consideration. In addition, it should be noted that the solution in explicit form for multidimensional systems is extremely rare, which was done in this work. And, it can also be added that the consideration of a problem in which the player does not know exactly the state of another player, but only knows that his state belongs to a fuzzy set, is an additional confirmation of the novelty of the problem under consideration. The model is intended for the computational core of the decision support system (DSS) in the tasks of optimizing investment strategies in the information security information security system. Computational experiments were carried out for the developed model.

Keywords: Information security systems · Investment · Optimal strategies · Decision support · Differential game · Simulation · Fuzzy set

1 Introduction

Any organization or company (hereinafter referred to as information security objects or OIS) possesses valuable information assets that require protection. Given the growing number of cyber-attacks recorded worldwide [1, 2], protecting against such attacks is

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 12–22, 2024.
https://doi.org/10.1007/978-3-031-50151-7_2

a pressing concern for almost all OIS. It is essential to prevent the leakage of valuable information through any channels from the OIS or its information system (IS). While information security (IS) management is increasingly recognizing the importance of protecting information assets, studies [3, 4] have shown that many information security specialists realize the current investment in IS is inadequate to counter all possible attacks on information assets. Moreover, a fixed security budget [5, 6] may not be sufficient for information security departments to address all threats and vulnerabilities in IS. Therefore, the process of investing in IS for OIS should be viewed as a continuous process. From a scientific research perspective, developing new models to manage continuous investment in information security systems is of interest. It is important to note that there are not only direct costs associated with ISS, but also indirect costs. While the direct costs for information security, such as hardware and software development and implementing security controls, may be clear to company management, the indirect costs are not as easily recognizable [7–9]. However, the opposing party in this game is not necessarily a singular entity. The players in the venture investment market for information security can also be considered as the opposing side. For instance, Forbes reports that in 2021, the global volume of venture investments in cybersecurity reached $21.8 billion, breaking the world investment record. The vast majority of these funds were raised by startups from the US and Israel. Additionally, according to the consulting company Gartner, the number one cyber threat in 2022 will be attacks on APIs, which are application programming interfaces. In response to this threat, two Israeli start-ups, Salt Security and Noname Security, have emerged in the API protection services and products market. Other innovative developments include Priori, an incident response platform that uses continuous monitoring of malicious activity known as BreachQuest. A Cylentium software product designed to hide the network and any device connected to it from the "eyes" of intruders. All of the above allows us to conclude that the investment strategy in the market of solutions in the field of information security largely involves creating a system of financial measures aimed at achieving long-term strategic goals related to information security solutions. Such goals, for example, could include ensuring a consistently high level of information security, which can be attained by combining both constant and variable parameters of the investment project.

2 Literature Review and Problem Formulation

Relying on the conclusions obtained by the authors in [8, 9], we can say that in order to assess investments in information security, it is advisable to use not only traditional methods and models [10], but also others that adequately reflect the essence of the problem. For example [12], points out that methods widely used in DSS are rarely suitable for synthesizing predictive estimates regarding the expediency of an investor's choice of an optimal strategy for investing in projects related to OIS IS. The game approach in the problems of choosing optimal investment strategies is also considered in [13, 14]. However, the analyzed models [14–16] based on game theory do not provide actual recommendations for investors. This also applies to situations where the optimal strategy for continuous financial investment in information security projects needs to be mathematically justified. It should be noted that the approaches presented in [16–18] do not allow for the full determination of an investor's optimal strategy in the OIS IS, and the task is complicated by the multifactorial nature of the information security products market. Investors may find it challenging to decide where to focus their attention, given that there are various solutions, such as protecting the OIS cloud infrastructure, WAF, monitoring OIS assets and networks, identity management, access control to confidential data, and more, as previously discussed. Information security innovations are becoming one of the primary mechanisms for ensuring maximum protection of information assets. Time is a critical factor in the innovation and investment development of the information security system for the OIS, even more so than money, as it is a more versatile tool in such investment projects. These circumstances determined the relevance of the development of new models designed to manage the continuous investment process in the OIS IS.

3 Goal

Goal: to develop a model for selecting a strategy for mutual continuous investment in information security solutions, considering the multifactorial nature of the process and presenting it in a fuzzy formulation. During this study, the following tasks will be accomplished: developing a decision support system model for selecting a strategy for mutual continuous investment in solutions related to the information security of the information security system, considering the multifactorial nature of the process and presenting it in a fuzzy formulation; conducting computational experiments to confirm the model's performance.

4 A Model for Choosing a Strategy for Mutual Continuous Investment in the Field of Information Security, Taking Into Account the Multifactorial Nature of the Problem in a Fuzzy Formulation

The developed model is a continuation of [13] and is based on solving a bilinear differential quality game with several terminal surfaces. The investment process in the complex area of information security highlights the challenge of allocating financial resources rationally to the most critical areas of protecting OBI information assets. There are no universal methods or models that can solve all existing problems, including those related to investment in information security. The aim of this study is to create a synergistic effect with existing methods and models widely presented in scientific literature to answer the question of which investment strategy in information security is more rational and what is the probability of an investor not losing their financial resources aimed at developing information security solutions. The task is formulated as follows: Two investors (players) manage a dynamic system in a multidimensional space formed by bilinear differential equations with dependent motions.

Terminal surfaces S_0, F_0 and sets of strategies U and V players are indicated. The first player The task (let's designate it as further $Inv1$) strives to bring the dynamical system to the terminal surface S_0 with the help of its control strategies. It does not matter to him how the second player (which we will denote as $Inv2$) will act. The task $Inv2$ is symmetrical.

We believe that at the moment of time $t = 0$ the players ($Inv1$ and $Inv2$) have allocated, respectively, $h(0) \in R_+^n$ and $f^{\xi}(0) \in R_+^n$ financial resources (hereinafter, we use the designation *FinR*) for investing in the corresponding IS. Players interact continuously. Such an interaction can be described as a bilinear differential game with fuzzy information. In contrast to a game with complete information, $Inv1$ doesnt know the initial state of $Inv2$. He only has access to information that the state of $Inv2$ belongs to a fuzzy set $\{F, m(.)\}$, where F is the subset R_+^n, $m(.)-$ is the function of the f^{ξ} state belonging to the set F, $m(f^{\xi}) \in [0, 1]$ $[0, 1]$ for $f^{\xi} \in F$. Further calculations in the work are presented from the $Inv1$ position. Thus, no assumptions about $Inv2$ awareness is made. In other words, $Inv2$ has any information. The interaction dynamics $Inv1$ and $Inv2$ is given by such a system of bilinear differential equations with dependent motions.

$$\begin{aligned} dh(t)/dt = -h(t) + B_1 \times h^+(t) + [(A_1 + R_1) - E] \times U(t) \times \\ \times B_1 \times h^+(t) - [(A_2 + R_2) - E] \times V(t) \times B_2 \times f^{+,\xi}(t); \\ df^{\xi}(t)/dt = -f^{\xi}(t) + B_2 \times f^{+,\xi}(t) + [(A_2 + R_2) - E] \times \\ \times B_2 \times f^{+,\xi}(t) - [(A_1 + R_1) - E] \times U(t) \times B_1 \times h^+(t); \end{aligned} \tag{1}$$

where $t \in [0, T]$, $T \in R_+$;

$h(t) \in R^n, f^{\xi}(t) \in R^n$, $U(t)$, $V(t)-$ square diagonal matrices of n order containing positive elements $u_i(t), v_i(t) \in [0, 1]$;

B_1, B_2—are the transformation matrices DF $Inv1$ and $Inv2$, which are square n order matrices with positive elements g_1^{ij}, g_2^{ij},, respectively;

A_1, R_1—diagonal matrices with positive entries. The elements correspond to $Inv2$ percentage fee for financial investments and $Inv2$ share of return on investment in relation to $Inv1$ investments in the information security solution;

A_2, R_2—diagonal matrices with positive entries. Likewise, the elements correspond to $Inv1$ percentage fee for financial investments and $Inv1$ share of return on investment in relation to $Inv2$ investments in information security solutions;

E—identity matrix;

$$x^+ = \begin{cases} x, x \geq 0; \\ 0, x \prec 0, x \in R; \end{cases}$$

The interaction of players ends if conditions (2) and (3) are met:

$$(h(t), f^{\xi}(t)) \in S_0, \text{ with certainty} \geq p_0 (0 \leq p_0 \leq 1) \tag{2}$$

$$(h(t), f^{\xi}(t)) \in F_0, \text{ with certainty } \geq \mathrm{p}_0 \tag{3}$$

where

$$S_0 = \bigcup\nolimits_{i=1}^{n} \{(h, f^{\xi}) : (h, f^{\xi}) \in R^{2n}, h \succ 0, f_i^{\xi} = 0\},$$

$$F_0 = \bigcup\nolimits_{i=1}^{n} (h, f^{\xi}) : (h, f^{\xi}) \in R^{2n}, f^{\xi} \succ 0, h_i = 0\},$$

We believe that if condition (2) is met, then the procedure for continuous investment in the information security solution is over. This means that *Inv*2 did not have enough *FinR* to continue the continuous investment procedures. At least this is true for one of the information security decisions considered by the player (with certainty $\geq p_0$).

If condition (3) is met, we also assume that the procedure for continuous investment of the IS solution is over. In other words, *Inv*1 did not have enough *FinR* to continue the procedures of continuous investment. Similarly, this is true for at least one of the IB decisions considered by the player (with certainty $\geq p_0$).

If conditions (2) and (3) are not met, then we assume that the procedures for continuous investment in information security solutions continue. The above process is considered within the framework of the scheme of a positional differential quality game with fuzzy information [19]. The task of the players (both *Inv*1 and *Inv*2) is to bring the dynamic system with the help of their control strategies to their terminal surface. Accordingly, the first player is on S_0 and the second is on F_0. Each player does not care how the other side acts.

Tasks for players are symmetrical. This means that two tasks are generated. Accordingly, the first from the point of view of the first player-ally. The second from the point of view of the second player-ally [13]. The article considers the problem from the point of view of the first player-ally. The symmetrical problem is not considered. The solution consists in constructing preference sets for players, i.e. *Inv*1 and *Inv*2. These preference sets are denoted as W_1 and W_2, respectively. In fact, the preference set of the player, for example, W_1 is the set of initial states of the first player and the uncertainty characteristics of the initial states of the second player. And the strategy, as mentioned earlier, is the formation by the player of a system of financial measures to achieve long-term strategic goals related to decisions in the field of information security of the OIS.

The first player (*Inv*1) in problem 1 is considered an ally player. Then the second player (*Inv*2) is considered the opposing player. In problem 2, the opposite is true. Then here the second player (*Inv*2) is the ally player, and *Inv*1- the enemy player. Players are interested in their investments in information security solutions.

Setting the procedure for the interaction of players c with the help of system (1) generates at each moment of time t a set of pairs of fuzzy sets $\{H_t, n_t(.)\} \times \{F_t, m_t(.)\}$. These fuzzy sets will reflect the processes of transition from the initial states of the players $(h(0), f^{\xi}(0))$ to subsequent states. Transition processes occur as a result of the application of control actions by the players.

We assume that *Inv*1 at each moment t $(t \in [0, T])$ has known states $h(\tau)$ for $\tau \leq t$. In this case, the following conditions are true:

$h(\tau) \geq 0$, if the reliability of states $n_\tau(h(\tau)) \geq p_0$ and $h(\tau) \notin R_+^n$; if the reliability of such states is $n_\tau(h(\tau)) \prec p_0$, and the values of the implementations of the *Inv*1$U(\tau)(\tau \leq t)$ strategy are known.

Then the following function can be defined:

$$F(.): X \to R_+, F(x) = \{\sup m(y), \text{для} \;\; y \leq x, x \in R_+^n\}. \tag{4}$$

Denote by Φ—the set of such functions. Then, through $T^* = [0, \infty]$—denote the time segment. The definition of a pure strategy $U(., ., .)$ of players was given in [13]. In fact, a pure strategy can be represented as a set of functions $u_i(., ., .) : T^* \times R^n_+ \times \Phi \to [0, 1]$, $(i = 1, ..., n)$.

We believe that *Inv*2 will choose a strategy $V(.)$ guided by any information about the *Inv*1 solution in the field of information security.

The first player (*Inv*1) aims to find the set of its initial states and the uncertainty characteristics of the initial states of the second player that satisfy the following condition A.

Condition A: if the game starts from such initial states, then *Inv*1, by choosing its strategy $U_*(.)$ can ensure the fulfillment of (2) at one of the t time points. At the same time, this strategy, chosen by *Inv*1, contributes to preventing the implementation of (3) at previous times. The set of such states is called the preference set *Inv*1-W_1. Therefore, *Inv*1 strategies, which have the above properties, can be considered optimal *Inv*1 strategies. The goal of the first player is to find the preference set and their corresponding strategies, which will fulfill condition (1). The game model formulated in this paper corresponds to the problem of decision making under conditions of fuzzy information, according to the classification of decision theory. The model is a bilinear quality differential game with several terminal surfaces.

Note that the search for preference sets *Inv*1 and its optimal strategy is associated with a set of parameters.

In order to describe preference sets *Inv*1, we introduce a number of notations and quantities.

Let's define a set $C(p_0) = \{c(0) : F(c(0)) \geq p_0\}$. For any $x \in R^n_+$ consider the set $L_x = \{z : z = l \times x, l \in R_+\}$. For any $x \in R^n_+$ consider the set $Q(x, p_0) = C(p_0) \cap L_x$.

Define a vector $\delta(x, p_0) : \delta(x, p_0) = \inf\{\delta^* : \delta^* \in Q(x, p_0)\}$.

Consider the set $\Delta(p_0) = \{\delta(p_0) : \exists x \in R^n_+ : \delta(p_0) = \delta(x, p_0)\}$.

Let us write down the conditions that make it possible to find a solution to the game. We are talking about the set of "preferences" W_1 and the *Inv*1 optimal strategy. These conditions will be written in terms of matrix inequalities: $(A_1+R_1)-E \succ 0$, $(A_2+R_2)-E \succ 0$; $(A_1+R_1)-E \succ 0$, $(A_2+R_2)-E \leq 0$; $(A_1+R_1)-E \leq 0$, $(A_2+R_2)-E \succ 0$; $(A_1+R_1)-E \leq 0$, $(A_2+R_2)-E \leq 0$ all other cases of ratios of data elements of matrices.

Let us introduce notation.

$$G^i = \sum_{j=1}^{n} [(A_2 + R_2) \times B_2]_{ij};$$

$$S^i = \sum_{\theta=1}^{n} \{[(A_1 + R_1) - E] \times B_1\}_{i\theta} / \{\sum_{j=1}^{n} \{[(A_1 + R_1) - E]_{ij}\}$$
$$\times \sum_{j=1}^{n} \{[(A_2 + R_2) - E] \times B_2\}_{\theta j}\};$$

$$G_1^i = \sum_{\theta=1}^{n} \{[(A_1 + R_1) - E] \times B_1\}_{i\theta} / \{\sum_{j=1}^{n} \{[(A_1 + R_1) - E]_{ij}\}$$

$$\times \sum_{j=1}^{n} \{[(A_1 + R_1)] \times B_1\}_{\theta j}\};$$

$$S_1^i = \sum_{j=1}^{n} \{[(A_1 + R_1) - E] \times B_1\}_{ij}; \;\; F^i = \sum_{j=1}^{n} [(A_1 + R_1) \times B_1]_{ij};$$

$$H^i = \sum_{\theta=1}^{n} \{[(A_2 + R_2) - E] \times B_2\}_{i\theta} / \{\sum_{j=1}^{n} \{[(A_2 + R_2) - E]_{ij}\}$$

$$\times \sum_{j=1}^{n} \{[(A_1 + R_1) - E] \times B_1\}_{\theta j}\};$$

$$H_1^i = \sum_{j=1}^{n} \{[(A_2 + R_2) - E] \times B_2\}_{ij};$$

$$F_1^i = \sum_{\theta=1}^{n} \{[(A_2 + R_2) - E] \times B_2\}_{i\theta} / \{\sum_{j=1}^{n} \{[(A_2 + R_2) - E]_{ij}\}$$

$$\times \sum_{j=1}^{n} \{[(A_2 + R_2)] \times B_2\}_{\theta j}\};$$

$$(q_*)_i = \frac{[G^i - G_1^i]}{[2 \times S_1^i]} + \sqrt{\{\frac{[G^i - G_1^i]}{[2 \times S_1^i]}\}^2 + [\frac{S^i}{S_1^i}]};$$

$$(\varphi_*)_i = \frac{[F^i - F_1^i]}{[2 \times H_1^i]} + \sqrt{\{\frac{[F^i - F_1^i]}{[2 \times H_1^i]}\}^2 + [\frac{H^i}{H_1^i}]};$$

The preference set W_1 for case (1) and within the framework of the above notation is defined as follows: $W_1 = \bigcup_{i=1}^{n} W_1^i; W_1 = \bigcup_{i=1}^{n} W_1^i;$

$$W_1^i = \bigcup_{i=1}^{n} \{(h(0), \delta(p_0)) : (h(0), \delta(p_0)) \in R_+^{2n}, (q_*)_i \times \delta_i(p_0) \prec$$

$$\sum_{\theta=1}^{n} \{[(A_1 + R_1) - E]_{i\theta} / \sum_{j=1}^{n} [(A_1 + R_1) - E]_{ij}\} \times h_\theta(0),$$

$$\forall i = 1, ..., n; \; (\varphi_*)_i \times h_i(0) \geq \sum_{\theta=1}^{n} \{[(A_2 + R_2) - E] \times B_2\}_{i\theta}$$

$$\times \delta_\theta(p_0) / \{\sum_{j=1}^{n} \{[(A_2 + R_2) - E] \times B_2\}_{ij}\};$$

The $Inv1$ optimal strategy would be a dependence of the form $U^*(t) = E.U^*(t) = E$.

Similarly, $Inv1$ optimal strategies and preference sets are found for cases (2–5).

In the same way, the solution of the problem is found from the side of the second player-ally.

Note that the variable $\delta(p_0)$ characterizing the degree of uncertainty of the initial states of the second player "participates" in the determination of the preference set. It «participates» in determining the preference set of the first player. This variable is an element of the set $\Delta(p_0)$.

5 Computational Experiment to Find Rational Investor Strategies

The computational experiment presented in this paper reflects a scenario where information security is focused on protecting a system comprising multiple complexes. In such a system, each complex must be protected as a priority, since any breach in the protection of one complex can lead to the entire system being compromised. For instance, if the software controlling the critical units of a nuclear power plant is damaged, it can lead to a system failure. In reality, complete awareness of the situation is rare, and hence it is crucial to develop recommendations for decision-making under incomplete information, where fuzzy information is a particular case. This highlights the significance of the problem formulation presented in this paper. To provide clarity of the control process, a non-standard way of illustrating the procedure was chosen, represented in the form of points in corresponding spaces. The authors used a program based on algorithms developed using the Bellman optimality principle to find the elements of the preference sets of the first player and their optimal strategies. As the approach presented in this paper is constructive and allows for the explicit formulation of optimal strategies, it reduces the complexity of calculations. Consequently, the use of less powerful computers is possible to find solutions to the problem under consideration, which is advantageous. The mathematical model presented in this work was implemented using the Python algorithmic language in the Anaconda development environment. The initial data used for the model were provided by several companies in Ukraine and Kazakhstan specializing in the development and implementation of their own information security solutions. The simulation results obtained using the developed software product are illustrated in Fig. 1. The graph corresponds to the results of the search for the preference set W_1 for the first investor for 4 variables.

6 Discussion of the Results of a Computational Experiment

Figure 1 shows a set of points in a four-dimensional space. The coordinate axes respectively characterize the following parameters: $H1(0)-Inv1$ financial resources (or $FinR$) allocated for investment in OIS IS projects. $H2(0)-Inv2$ financial resources (or $FinR$) allocated for investment in OIS IS projects. $Delta0(P0)-$ the degree of reliability of the procedure for investing in OIS information security (with a given level of reliability) for the first variable. $Delta1(P0)-$ the degree of reliability of the procedure for investing in OIS information security (with a given level of reliability) for the second variable. For each $FinR$ point that corresponds to the first player, we have the values of the variables $Delta0(P0)$, $Delta1(P0)$.

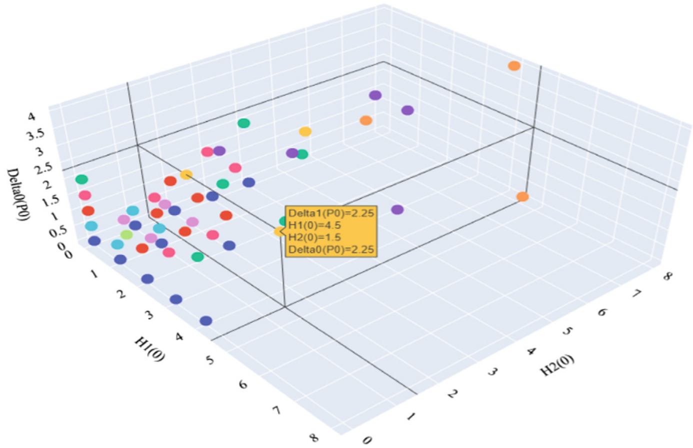


Fig. 1. Dependence of the preference set W_1 for the first investor for 4 variables

There may be several of these values. Some of these values will correspond to a set that guarantees the continuation of the procedures for investing in information security. The other part corresponds to a situation in which *Inv*2 cannot continue investing with a given level of certainty. Therefore, choosing the minimum values for each component from the available values, we can obtain a preference set *Inv*1.

This statement is true under such a condition—all values of the variable *Inv*2, which characterize the reliability of the procedure for investing in information security, will be less than the previously given minimum values of the "reliability" variables. The graphical interpretation of the processes of managing the investment procedure in information security shown in Fig. 1 corresponds to the model of continuous investment. *Inv*1 can use *FinR*, which are defined by the given sets of financial resources. In turn, these sets are determined by the choice of specific investment programs in information security solutions.

For example, these can be the solutions mentioned above for protecting the cloud infrastructure of the OIS, WAF, monitoring assets and the OIS network, identity management, access control to confidential data, and much more. Note that the choice shown in Fig. 1 of the method of illustrating the preference set *Inv*1 is not accidental. This method is convenient when it is necessary to graphically illustrate situations in spaces of greater dimensions than three. In our opinion, this is a positive moment.

The model proposed in this paper does not claim to provide all the answers to investors during the implementation of information security projects. The actual list of variables is much larger than what was considered in the problem formulation. Our aim was not to cover all possible project options since other methods and models, such as T. Saaty's traditional analytical hierarchy process [20–27], and SWOT analysis [22, 23], exist for that purpose. However, we believe that in managing the investment process in information security, a synergy of various methods and models is necessary to obtain the desired answer to the question of which investment strategy is more rational and what the probability is for the investor to avoid financial losses when developing solutions in the field of information security.

7 Conclusion

The paper proposes a model for managing continuous investment in information security systems (ISS) of information security objects (OIS). The scientific novelty of the model lies in considering continuous investment in the protection of information security, taking into account multifactoriality in a fuzzy formulation. Additionally, the paper describes a new class of bilinear differential games in a fuzzy formulation. The model is intended for the computational core of the decision support system (DSS) to optimize investment strategies in the information security system. The mathematical model was implemented in the Python algorithmic language in the Anaconda development environment. Computational experiments were conducted, and a graphical interpretation of finding the preference set for the first investor for four variables was obtained. The results of this work can help rationally use available financial resources to address society's pressing problems. Ongoing research is focused on creating a decision support system based on an ensemble of models using the apparatus of bilinear differential games. Such a DSS will provide investors in information security systems with more accurate predictive estimates for investment projects of interest.

References

1. Lallie, H.S., et al.: Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Comput. Secur. **105**, 102248 (2021)
2. Upadhyay, N.K., Rathee, M.: Cyber security in the age of Covid-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Med., Law Soc. **15**(1), 89–106 (2022)
3. Hallman, R.A., et al.: Determining a return on investment for cybersecurity technologies in networked critical infrastructures. Int. J. Organ. Collect. Intell. (IJOCI) **11**(2), 91–112 (2021)
4. Kalderemidis, I., Farao, A., Bountakas, P., Panda, S., Xenakis, C.: GTM: game theoretic methodology for optimal cybersecurity defending strategies and investments. In: Proceedings of the 17th International Conference on Availability, Reliability and Security, pp. 1–9 (2022)
5. Garg, V.: Covenants without the sword: market incentives for cybersecurity investment. Available at SSRN 3896578 (2021)
6. Mazzoccoli, A., Naldi, M.: Optimal investment in cyber-security under cyber insurance for a multi-branch firm. Risks **9**(1), 24 (2021)
7. Sawik, T., Sawik, B.: A rough cut cybersecurity investment using portfolio of security controls with maximum cybersecurity value. Int. J. Prod. Res. **60**(21), 6556–6572 (2022)
8. Kuzlu, M., Fair, C., Guler, O.: Role of artificial intelligence in the Internet of Things (IoT) cybersecurity. Disc. Internet of things **1**(1), 1–14 (2021)
9. Choo, K.K.R., Gai, K., Chiaraviglio, L., Yang, Q.: A multidisciplinary approach to Internet of Things (IoT) cybersecurity and risk management. Comput. Secur. **102**, 102136 (2021)
10. Holovkin, B.M., Tavolzhanskyi, O.V., Lysodyed, O.V.: Corruption as a cybersecurity threat in conditions of the new world's order. Ling. Cult. Rev. **5**(S3), 499–512 (2021)
11. Mukhopadhyay, I.: Cyber threats landscape overview under the new normal. In: ICT Analysis and Applications, pp. 729–736. Springer, Singapore (2022)

12. Akhmetov, B.B., Lakhno, V.A., Akhmetov, B.S., Malyukov, V.P.: The choice of protection strategies during the bilinear quality game on cyber security financing. Bull. Nat. Acad. Sci. Republic of Kazakhstan **3**, 6–14 (2018)
13. Lakhno, V., Malyukov, V., Gerasymchuk, N., et al.: Development of the decision making support system to control a procedure of financial investment. East.-Europ. J. Enterp. Techn. **6**(3), 24–41 (2017)
14. Smit, H.T., Trigeorgis, L.: Flexibility and games in strategic investment (2015)
15. Arasteh, A.: Considering the investment decisions with real options games approach. Renew. Sustain. Energy Rev. **72**, 1282–1294 (2017)
16. Wang, Y., Gao, W., Qian, F., Li, Y.: Evaluation of economic benefits of virtual power plant between demand and plant sides based on cooperative game theory. Energy Convers. Manage. **238**, 114180 (2021)
17. Sahin, B., Yazir, D., Soylu, A., Yip, T.L.: Improved fuzzy AHP based game-theoretic model for shipyard selection. Ocean Eng. **233**, 109060 (2021)
18. Meng, H., Liu, X., Xing, J., Zio, E.: A method for economic evaluation of predictive maintenance technologies by integrating system dynamics and evolutionary game modelling. Reliab. Eng. Syst. Saf. **222**, 108424 (2022)
19. Lakhno, V., Malyukov, V., Kasatkin, D., Chubaieskyi, V., Rzaieva, S., Rzaiev, D.: Continuous investing in advanced fuzzy technologies for smart city. Lect. Notes Data Eng. Comm. Tech. **142**, 313–327 (2023)
20. Polyakov, V.A., Fomicheva, I.V.: Express assessment of the investment attractiveness and competitiveness of regional territories. J. Reg. Int. Comp. **3**(2), 24 (2022)
21. Vovk, O., Kravchenko, M., Popelo, O., Tulchynska, S., Derhaliuk, M.: Modeling the choice of the innovation and investment strategy for the implementation of modernization potential. Trans. Syst. Cont. **16**, 430–438 (2021)
22. Yao, D., de Soto, B.G.: A preliminary SWOT evaluation for the applications of ML to cyber risk analysis in the construction industry. In: IOP Conference Series: Materials Science and Engineering, Vol. 1218, No. 1, p. 012017. IOP Publishing (2022)
23. Emer, A., Unterhofer, M., Rauch, E.: A cybersecurity assessment model for small and medium-sized enterprises. IEEE Eng. Manage. Rev. **49**(2), 98–109 (2021)
24. Gromov, D., Gromova, E.: On a class of hybrid differential games. Dynam. Games Appl. **7**(2), 266–288 (2017)
25. Batukhtin, V.D., Krasovskii, N.N.: Extremal control in a nonlinear positional differential game. Akademiia Nauk SSSR, Izvestiia, Tekhnicheskaia Kibernetika, 55–63 (1973)
26. Bebeshko, B., Malyukov, V., Lakhno, M., Skladannyi P., Sokolov V., Shevchenko S., Zhumadilova, M.: Application of game theory, fuzzy logic and neural networks for assessing risks and forecasting rates of digital currency J. Theor. Appl. Info. Tech. **100**(24) (2022). http://www.jatit.org/volumes/Vol100No24/15Vol100No24.pdf
27. Bebeshko, B., Khorolska, K., Desiatko, A.: Analysis and modeling of price changes on the exchange market based on structural market data. In: 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), pp. 151–156 (2021). https://doi.org/10.1109/PICST54195.2021.9772208

Modelling and Forecasting Foreign Direct Investment: A Comparative Application of Machine Learning Based Evolutionary Algorithms Hybrid Models

Mogari Ishmael Rapoo[1], Martin Chanza[2], and Elias Munapo[2](✉)

[1] Department of Mathematics & Physics, Cape Peninsula University of Technology, Bellville Campus, Cape Town, South Africa
rapoom@cput.ac.za

[2] Department of Business Statistics and Operations Research, School of Economic Sciences, North West University, Mafikeng Campus, Mahikeng, South Africa
Elias.Munapo@nwu.ac.za

Abstract. The current study sought to determine whether genetic algorithm can improve forecasting performance and accuracy of machine learning models (ANN, SVR and LSSVR) in a hybrid setting for modelling and forecasting foreign direct investment. The study employed artificial neural network (ANN), support vector regression (SVR) and least squares support vector regression (LSSVR) as benchmark models to determine the forecasting performance of both artificial neural networks based genetic algorithm (ANN-GA) and support vector regression based genetic algorithm (SVR-GA) models respectively. Monthly time series data of foreign direct investment, exchange rate, inflation rate and gross domestic product from Jan 1970 to June 2019 were employed as time series data. The results showed that ANN had a better forecasting accuracy than both SVR and LSSVR respectively, and in the hybrid models SVR-GA had a better forecasting performance than ANN-GA. In the overall it was found that genetic algorithm does not improve the forecasting performance and accuracy of machine learning models as show by MSE, RMSE, MAE, MAPE, and MASE. The study recommends that other evolutionary algorithms be adopted and used to optimize hyper-parameters of machine learning models.

Keywords: Error measurements · Evolutionary algorithm(s) · Foreign direct investment hybrid models · Modelling and forecasting · Machine learning

1 Introduction

According to Khashei and Bijari (2011), considerable amount of attention has been given to time series forecasting in application to the variety of areas. Chujai et al. (2013) define time series as a group of data collected over time according to a continuous period of time or an order of historical data similar to an observation. Various models have been employed in the past for modelling and forecasting time series data. These

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 23–35, 2024.
https://doi.org/10.1007/978-3-031-50151-7_3

conventional models include regression analysis, classical decomposition methods, Box and Jenkins methods and smoothing techniques. However, Wheelwright et al. (1998) state that several factors are questioned in consideration of an appropriate model to be used.

The most extensively applied model for time series modelling and forecasting has been the Box-Jenkins auto-regressive integrated moving average (ARIMA) model. This model has been applied successfully in various disciplines and particularly economic time series forecasting, such as for exchange rate forecasting (Nyoni 2018; Nwankwo 2014), and for dependence between successive times and failure (Walls and Bendell 1987). A linear structure is usually assumed among the time series values in the existing studies. However, Zhiqiang et al. (2013) states that in the real-world phenomenon, linear relationship is not factual as there exists a nonlinear relationship. Therefore, this presents a limitation to the ARIMA model since it is a linear model.

According to Chaudhry et al. (2000), knowledge-based decision support is provided through the development of applied artificial intelligence system amongst others artificial neural networks (ANNs). These machine learning models are data driven and can approximate any function because of their exceptional properties. The other advantage of these models is that they are nonparametric, meaning no or fewer prior assumptions are needed in the model building process. These models are also nonlinear in nature. These models include, but not limited to, artificial neural networks (ANNs), support vector regression (SVR), adaptive neuro-fuzzy inference systems (ANFIS) and neural networks autoregressive with exogenous inputs (NNARX).

In recent years, these models have been efficiently carried out in various research areas, such as for rainfall run-off simulation (Talei et al. 2010; Nourani et al. 2009); electricity forecasting (Kaytez et al. 2015) and financial time series forecasting (Tay and Cao 2001). These models have obtained exceptional results in many research areas.

Artificial intelligence has a branch which has been developed based on the theory of evolution. This branch is computational methods called evolutionary algorithms. The Darwin's evolution and genetics was used in the development of computational methods called evolutionary algorithms, amongst others, there is genetic algorithm. These methods are used to search for optimum solutions and are applied in a hybrid setting together with the various machine learning models.

The prime objective of this approach is that the algorithm can help in the optimal selection of the hyper-parameters of machine learning models. Furthermore, it has been found that machine learning models depend heavily on the right choices of the hyper-parameters to improve forecasting accuracy. Various scholars have employed these hybrid models for time series forecasting, for example, in water quality prediction (Liu et al. 2013); system reliability (Chen 2007); tourism demand forecasting (Chen and Wang 2007); long span suspension bridges (Cheng 2010) and electric load forecasting (Hong 2009). Thereby, it was found that these algorithms can improve the forecasting accuracy of machine learning models.

Therefore, this research seeks to investigate the application of machine learning based evolutionary algorithms hybrid models in terms of their forecasting efficiency and performance of modelling and forecasting foreign direct investment. Furthermore, to determine if the evolutionary algorithms can improve the forecasting accuracy of the included machine learning models in modelling and forecasting foreign direct investment, the study uses traditional machine learning models as benchmark models.

The study is organized as follows: Sect. 2 provides a review of the literature, Sect. 3 explains the methodology, Sect. 4 presents the empirical findings, and Sects. 5 provides a conclusion and recommendations respectively.

2 Literature Review

According to Shalev-Shwartz and Ben-David (2014), "machine learning is one of computer science's fastest-growing fields, with a wide range of applications." Automated pattern recognition in data is referred to as machine learning (ML). Machine learning has recently been widely used for practically any application that requires information extraction from large data sets. Machine learning is also used in many different industries, including scientific subjects like bioinformatics, medicine, and astronomy. Giving programs the capacity to learn and adapt is the goal of machine learning tools or algorithms.

In the numerous models of artificial neural networks (ANNs) developed, the mostly approved and applied model is the feed-forward multilayer network or the multilayer perceptron (MLP). This model has been adopted and applied for forecasting purposes. This model is composed of several layers which are the input layer, output layer and in between is one or several hidden layers. Except in the input layer, the other layers have an activation function which helps in the calculation of the weights of the model.

In the study of Rapoo and Xaba (2017), the forecasting performance of ARIMA and ANN was investigated using South African rand and United States dollar (ZAR/USD) exchange rate series. The study employed mean squared error (MSE) and mean absolute error (MAE) as error metrics to compare the forecasting performance of the models. Based on the results, it was proven that ARIMA model had a superior performance to that of ANN in forecasting ZAR/USD exchange rate.

Barapatre et al. (2018) applied in their study multilayer feed-forward network trained with back-propagation as this model was also applied in the study of Bing et al. (2012) to predict stock market price. The results demonstrated that there is no other model in predicting the direction of changes of the value of stock better than multilayer feed-forward network trained with back-propagation algorithm.

Ghosh et al. (2019) in their study found out that response surface methodology (RSM) coupled with particle swarm optimization optimize the cutting conditions that lead to minimum surface roughness better than ANN coupled with genetic algorithm.

In their study, Orjuela-Canon et al. (2022) used three neural network models: a radial basis function for time series forecasting, a recurrent neural network, and a nonlinear autoregressive model. These models were trained using data on reported cases obtained from the national vigilance institution in Columbia. Based on the results' mean average percentage error, it can be seen that the traditional method-based model performs better than connectionist models.

For the in-sample forecasting of portfolio inflows, Rapoo et al. (2020) analysed the forecasting efficiency of support vector regression, artificial neural network, and VAR. Models were applied to series of gross domestic product (GDP), inflation-linked bonds, currency rates, and portfolio inflows covering the time period from 1 March 2004 to 1 February 2016. It was discovered that SVR performs forecasting more accurately than ANN models. Additionally, 9% of the variation in the factors that influence the allocation of PIs into South Africa as determined by structural vector auto-regressive (SVAR) was explained by push factors, whereas 69% of the variation was explained by pull factors.

For the prediction of the shear strength capacity of medium- to ultra-high strength concrete beams with longitudinal reinforcement and vertical stirrups, Jiang and Liang (2021) used a hybrid model of support vector regression based genetic algorithm (SVR-GA). Traditional SVR, a multiple linear regression model, and ACI-318 were contrasted with the suggested model. The proposed model outperformed the other models in the investigation, according to the coefficient of determination, root mean squared error, and mean absolute error.

A hybrid AI model using genetic algorithm-based support vector regression was used by Xu et al. (2021) to predict the electric energy consumption of crude oil pipelines. Three crude oil pipelines in China were the subject of the data used. The outcomes demonstrate that the suggested genetic algorithm-based support vector regression is the most effective model since it increases prediction accuracy when compared to to the state-of-the-art models.

The study employed support vector regression with genetic algorithm to prepare groundwater springs potentiality for the Jerash and Ajloun region, Jordan. The results showed that the desired precision level and major steps in modelling was achieved due to optimal values of kernels in SVR selection of optimal features (Al-Fugara et al. 2022).

By using a new multi-objective least squares support vector machine model with mixture kernels, Zhu et al. (2022) predicted asset prices. Also, the least squares support vector machine's mixture kernels' optimal model selection was synchronously searched via particle swarm optimization. The outcomes demonstrated that the suggested model can outperform other well-known models in terms of forecasting accuracy and trading performance.

The study used a short-term power load forecasting method based on AI algorithm known as IGA-LSSVM to develop a load forecasting model. Temperature, load, weather condition, working and holiday days are the input variables employed in the prediction model, and the output was load value. The information was obtained from a Yunnan province city. According to the study's findings, the proposed model was superior to others in terms of short-term power load prediction (Bao-De et al. 2021).

The study made a contribution to the field by combining the least squares support vector machine with the multi-objective sine cosine optimisation technique for forecasting electricity consumption. The findings demonstrated that, in the instance of Australia, the suggested model ensures high accuracy and great stability (Li et al. 2020).

By using ANFIS, ANN, and RMS models in a comparative study to model and optimize the substrate treatment process in biogas energy production from yam peel substrates via anaerobic digestion, the study made a significant contribution to the literature. According to the findings, ANFIS model performs just a little bit better when simulating and modeling anaerobic digestion (Onu et al. 2022).

In order to optimize the experimental response of total petroleum hydrocarbon (TPH) polluted soil, Ani and Agu (2022) used both ANFIS and ANN. For purposes of comparison and model performance evaluation, the study used RMSE, MSE, and MAE. The findings showed that the ANFIS model outperformed the ANN model in terms of model performance. It also demonstrates how much superior the ANFIS prediction capacity was than the ANN model.

3 Methodology

The study which is quantitative in nature uses monthly time series data of foreign direct investment, exchange rate, real GDP and inflation rate from the period 1970 to 2019 sourced from World Bank and South African Reserve Bank respectively. Furthermore, reported will be explanatory data analysis and the series is/(are) also tested for nonlinearity. The tests which are employed to assess nonlinearity are BDS test and CUSUM test.

Three supervised nonlinear machine learning models of artificial neural network (ANNs), support vector regression (SVR), and least square support vector regression (LSSVR) and one evolutionary algorithm (genetic algorithm) are employed to model and forecast foreign direct investments. Whereby, the evolutionary algorithm technique (genetic algorithm) is employed to optimize the hyper-parameters of both artificial neural network (ANN-GA) and support vector regression (SVR-GA) and to improve forecasting accuracy of the models. A brief description of the models is also shown:

Artificial neural network (ANN): ANNs are high performance, non-linear analytical tools that are capable of establishing a relationship between the dependent and independent variables without prior knowledge of the system. They are also nonparametric in nature. Multilayer perceptron (MLP) is a commonly used ANN approach for solving problems. The following structure shows the architecture of the MLP model (Fig. 1).

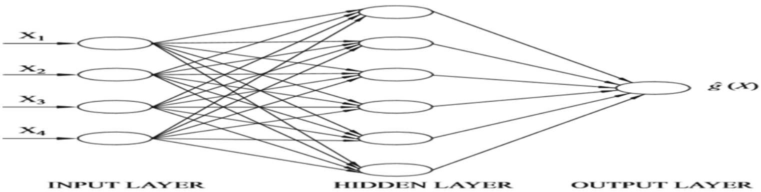


Fig. 1. Structure of Neural Network. *Source* Cheng (2010)

There are three layers in the structure of the neural network: input layer (independent variables), output layer (dependent variables) and one or more hidden layers. These layers have transfer functions. In general one hidden layer with sufficient neurons is good for solving problems, however this will depend on the complexity of the problem. Input and output layers will adopt purline which is a linear transfer function and in the hidden layer a nonlinear transfer function of sigmoid transfer function is employed (Karimi and Dastranj 2014). Hyper-parameters that need to be optimized so that the models forecasting performance may improve are number of neurons in the hidden layer, weights and bias.

Support vector regression (SVR): the model support vector machine (SVM) was initially intended to address classification problems (Vapnik 2013). Drucker et al. (1997) proposed SVR model which is the version of SVM. Support vector regression follows a principle of structural risk minimization which minimizes an upper bound of the generalization error than the prediction error on the training set (the principle of empirical risk). The linear regression function (in the feature space) using mathematical notation, is approximated using the following expression: $y = f(x) = \omega^T \varphi(x_i) = b$, $\varphi : R^n \rightarrow F, \omega \in F$, given that ω and b are coefficients, $\varphi(x_i)$ denotes the high dimensional feature space which is nonlinearly mapped from the input space x. The study will adopt the Gaussian function as the kernel function. Smola and Schölkopf (1998) stated that Gaussian function yields better prediction accuracy. Hyper-parameters that must be optimized are regularisation parameter C which determine the trade-off cost between minimizing the training error and minimizing the models complexity, bandwidth of the kernel function σ^2 which represents the variance of the Gaussian kernel function and the tube size of the ε-insensitive loss function (ε) which is equivalent to approximation accuracy placed on the training data points.

Least square support vector regression (LSSVR): least square support vector machine is an alternative of the standard support vector machine and it was improved by (Suykens et al. 2002). Least squares loss function is employed by LSSVM to construct the optimization problem based on the equality constraints. LSSVM is used commonly for optimal control, classification and regression problem (Suykens et al. 2002). The LSSVR technique is to approximate obscure function by using given a sample of training data series. The gression function can be formulated as feature space representation. Given by the follows mathematician notation: $y = f(x) = \omega^T \varphi(x_i) = b$, $\varphi : R^n \rightarrow F, \omega \in F$, given that ω and b are coefficients, $\varphi(x_i)$ denotes the high dimensional feature space which is nonlinearly mapped from the input space x. The current study will adopt radial basis function as a kernel function. LSSVR has equal constraints unlike the standard SVR. Hyper-parameters to be optimised are $Cand\sigma$.

Genetic algorithm (GA): genetic algorithm is a stochastic search technique that can be employed to search large and complicated spaces and the ideas of natural genetics and evolutionary principles are employed by GA. GA has two types real and binary coded GA respectively. The study will adopt real value GA to optimize the hyper-parameters of the machine learning models because it is straight forward, fast and efficient. Moreover, the study employed mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and mean absolute scaled error (MASE) and to assess forecasting performance and accuracy of the models.

4 Empirical Results

In this section both the preliminary data analysis and the main analysis are undertaken with the sole objective of addressing the aim of the study. This section will follow the methodology of this research.

4.1 Preliminary Data Analysis Results

The study reports on the time series plot of the four variables used in the study. It is imperative that the plots of the variables is/(are) done so that the characteristics of the variables can be apparent and easily to understand. Furthermore, Fig. 2 shows the plot of foreign direct investment (FDI), gross domestic product (GDP), inflation rate and exchange rate time series data from January 1970 to June 2019.

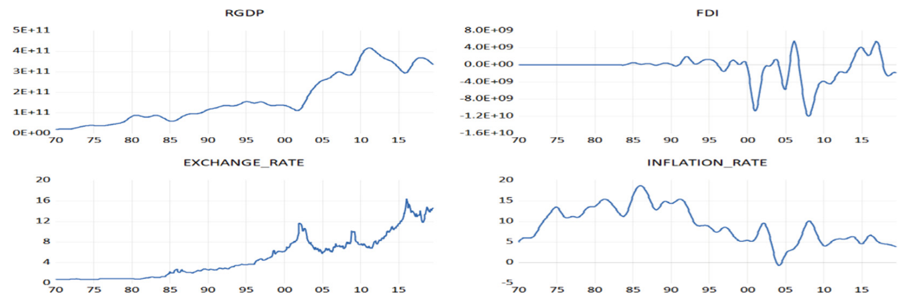


Fig. 2. Time Series plot of FDI, GDP, Exchange Rate, and Inflation Rate from January 1970 to June 2019

Time series plot of these four time series variables shows that FDI from January 1970 to June 2019 has been having a fairly equilibrium investment into South Africa. At some instances, it was having a negative investment into the country, meaning investors during this time were not investing in the form of FDI investment into South Africa. Moreover, inflation rate has also been constant on the month-to-month basis, with the aim of keeping it around the range of 3% and 6%. South African gross domestic product, on the other hand, has been steadily showing an upward trend which then becomes evident that there has been a good output from the country. ZAR/USD exchange rate presents a weak rand against US dollar, as the plot of the series shows an upward trend.

The following are the results of the nonlinearity test of both BDS test and CUSUM test of stability. These tests are employed to test whether a time series variable is linear or nonlinear in nature.

Table 1 reports on the BDS test results of nonlinearity. Based on the results, it is reported that the null hypothesis of linearity is rejected for foreign direct investment, gross domestic product, inflation rate and exchange rate in favour of nonlinearity, since for all the dimensions of BDS tests the probability values associated with the test are all 0.0000 for all the mentioned variables. This implies that FDI, GDP, inflation rate and exchange rate are all nonlinear in nature. Moreover, CUSUM test of stability is used to support the results of BDS test of nonlinearity. Based on CUSUM stability test results, Fig. 3 is shown to be unstable as the blue line on the graph is outside the critical red line on the graph, this then confirms the results of BDS test that the series of FDI, GDP, inflation rate, and exchange rate are inherently nonlinear.

The study trained twenty-four models of artificial neural network (ANN) and the best model took 9040 steps for it to converge with the minimum error value of 0.164947

Table 1. BDS nonlinearity test results

Variable(s)	Dimension	BDS statistic	Std. error	Z-statistic	P-value
Foreign direct investment	2	0.192	0.005	35.965	0.000
	3	0.322	0.009	37.798	0.000
	4	0.408	0.010	40.044	0.000
	5	0.464	0.011	43.404	0.000
	6	0.499	0.010	48.036	0.000
Gross domestic product	2	0.205	0.002	84.841	0.000
	3	0.348	0.004	90.889	0.000
	4	0.448	0.005	98.513	0.000
	5	0.517	0.005	109.568	0.000
	6	0.565	0.005	124.698	0.000
Inflation rate	2	0.201	0.002	103.054	0.000
	3	0.339	0.003	110.198	0.000
	4	0.433	0.003	119.160	0.000
	5	0.497	0.004	132.129	0.000
	6	0.539	0.004	149.812	0.000
Exchange rate	2	0.196	0.003	72.340	0.000
	3	0.331	0.004	77.398	0.000
	4	0.424	0.005	83.771	0.000
	5	0.489	0.005	93.027	0.000
	6	0.534	0.005	105.872	0.000

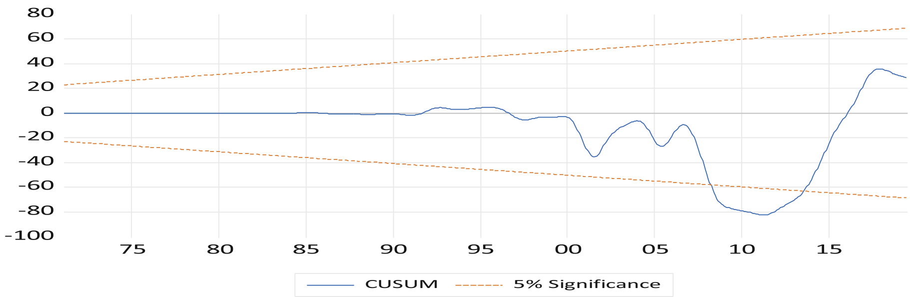


Fig. 3. CUSUM test of stability results

which was the smallest error amongst the twenty-four trained artificial neural network (ANN) models.

The chosen model of artificial neural network from the twenty-four models trained had three input nodes as this study employed three (exchange rate, inflation rate and gross domestic product) independent variables to foreign direct investment (FDI), three hidden layers with seven, six and five nodes per hidden layer. Furthermore, the study employed logistic activation function in the hidden layer of the network. This particular chosen model was used to test its generalization powers and the results will be reported in Table 2.

The study trained the initial model of SVR machine learning model and after training the model, root mean squared error (RMSE) was used as the measure of accuracy for the trained model and it was shown that the initial trained model had 0.1090777 RMSE value. Moreover, in the study, the optimal search of hyper-parameters of SVR model was undertaken to ensures an improvement of accuracy of the initial trained SVR model. In improving the accuracy of the initially trained model, the SVR model was tuned. This allowed the study to set an interval of tuned hyper-parameters of the study which are epsilon (ε) and cost (C) to $seq(0, 1, 0.1)$ and $(1 : 100)$ respectively. Thus, the optimal hyper-parameter values of both epsilon and cost chosen are 0.4 and 100 respectively with a best performance of the model at 0.0050630.

From the chosen tuned model of SVR, the study trained 1100 models of SVR and the best model of SVR which was chosen had a root mean squared error (RMSE) of 0.0449601 which shows an improvement of accuracy from the initially trained SVR model which had an RMSE value of 0.1090777. The overall chosen hyper-parameter values of epsilon, cost and gamma for the chosen best model of SVR are 0.4, 100 and 0.333 respectively, with radial basis function used as a kernel function and the model having 99 support vectors.

In the context of the current study, LSSVR model is employed to model and forecast foreign direct investment (FDI) using the radial basis function as a kernel function. The LSSVR model was trained using the method of trial and error with different values for the gamma hyper-parameter.

The performance of the LSSVR model significantly improved in terms of forecasting performance when the gamma value was set to be 0.20 instead of the 0.01 default value. This chosen model of LSSVR was chosen from a total of 63 LSSVR trained models and the model with hyper-parameter value of gamma which was 0.20 was the best trained model. The model showed a significant improvement in terms of generalisation.

The appropriate hyper-parameter optimisation of ANN model using GA was based on back-propagation algorithm. The hybrid model of genetic algorithm (GA) based artificial neural network (ANN) was executed using the TPOT in python environment. The optimal solution of the model hyper-parameters using GA was obtained after 50 iterations. The optimal hyper-parameters chosen by genetic algorithm are three hidden layer sizes with 50, 100, 50 as neurons per layer; hyperbolic tangent as activation function; alpha with the optimal value of 0.0001; and learning rate used is adaptive.

After the genetic algorithm (GA) was used to search for the above optimal parameters of ANN model, the ANN optimal model was built and trained on the train data. The *RMSE* of train data used in the study was at its minimum value which was found to be 0.1001. Furthermore, the study tested the hybrid model of GA-ANN against the unseen data of test data. The error measurements found on the test data showed that the model had a fairly good forecasting and generalisation of the test data.

The study, in terms of the process of searching for optimal hyper-parameters, was operated with 50 population size and the overall optimal fitness was reached at generations 10. Thus, the individuals at generation 10 produced the following optimal hyper-parameters of the model SVR, which are as follows; cost is $C = 5.807201$, gamma is $\sigma^2 = 1.660981$ and epsilon is $\varepsilon = 0.1818941$. The other parameters of GA are crossover probability value of 0.8, elitism value of 2 and mutation probability value of 0.1. The maximum iterations for which the genetic algorithm was trained is 10 with a fitness function value of -0.001575.

After the genetic algorithm (GA) was used to search for the above optimal parameters of the model SVR, the SVR optimal model was built and trained on the train data. The *RMSE* of train data used in the study was at its minimum value which was found to be 0.034445. Furthermore, the study tested the hybrid model of GA-SVR against the unseen data of test data. The error measurements found on the test data showed that the model had a good forecasting and generalisation of the test data.

And based on the overall comparison of the forecasting performance and accuracy of the models, Table 2 shows the results of various error measurements used to assess the performance of the models.

Table 2. Forecasting performance of traditional machine learning models and hybrid machine learning models

Models	Performance criterion				
	MSE	RMSE	MAE	MAPE	MASE
ANN	0.00180	0.04248	0.02817	5.67016	0.73075
SVR	0.00225	0.04743	0.03199	9.57313	0.082976
LSSVR	0.02527	0.15898	0.10125	15.66068	2.62555
Machine learning based genetic algorithm hybrid models					
Hybrid models	Performance criterion				
	MSE	RMSE	MAE	MAPE	MASE
ANN-GA	0.86663	0.93093	0.69106	392.35030	17.92044
SVR-GA	0.01437	0.11987	0.07797	42.07455	2.02197

Results shown in the above table starting with traditional machine learning models, it is evident that based on the error measurements values reported, artificial neural network (ANN) has the best forecasting performance and accuracy as compared to support vector regression (SVR) and least squares support vector regression (LSSVR) models respectively. Furthermore, it is also shown that in the hybrid models SVR-GA model had a better forecasting performance as compared to ANN-GA hybrid model. The overall results shows that the best overall performing model is the artificial neural network (ANN). It is also further shown that genetic algorithm (GA) did not improve the forecasting performance of both artificial neural network (ANN) and support vector regression (SVR).

5 Conclusion and Recommendations

A monthly time series covering the period of January 1970 to June 2019 for the variables foreign direct investment, exchange rate, inflation rate and gross domestic product was modelled using ANN, SVR, LSSVR, ANN-GA and SVR-GA respectively. The series showed that it is nonlinear in nature as both BDS and CUSUM test of stability were employed to test for linearity of the series respectively.

The study employed several error measurements which are MSE, RMSE, MAE, MAPE and MASE for assessing the forecasting performance of the models included and it was found that in the overall ANN was the best performing model as compared to SVR, LSSVR, ANN-GA and SVR-GA. Furthermore, it was also found out that genetic algorithm (GA) did not improve the forecasting performance of artificial neural network (ANN) and support vector regression (SVR) respectively. The study contributed to the literature by showing that for foreign direct investment genetic algorithm does not improve forecasting performance of the included machine learning models.

The study recommends that more machine learning models and other evolutionary algorithm techniques to be adopted and employed to model and forecast foreign direct investment in the future.

References

Al-Fugara, A.K., Ahmadlou, M., Al-Shabeeb, A.R., Alayyash, S., Al-Amoush, H., Al-Adamat, R.: Spatial mapping of groundwater springs potentiality using grid search-based and genetic algorithm-based support vector regression. Geocarto Int. **37**, 284–303 (2022)

Ani, K.A., Agu, C.M.: Predictive comparison and assessment of ANFIS and ANN, as efficient tools in modeling degradation of total petroleum hydrocarbon (TPH). Cleaner Waste Syst., 100052 (2022)

Bao-De, L., Xin-Yang, Z., Mei, Z., Hui, L., Guang-Qian, L.: Improved genetic algorithm-based research on optimization of least square support vector machines: an application of load forecasting. Soft. Comput. **25**(18), 11997–12005 (2021). https://doi.org/10.1007/s00500-021-05674-9

Barapatre, O., Tete, E., Sahu, C.L., Kumar, D., Kshatriya, H.: Stock price prediction using artificial neural network. Int. J. Adv. Res. Ideas Innov. Technol **4**, 916–922 (2018)

Bing, Y., Hao, J.K., Zhang, S.C.: Stock market prediction using artificial neural networks. Adv. Eng. Forum, Trans. Tech. Publ., 1055–1060 (2012)

Chaudhry, S.S., Varano, M.W., Xu, L.: Systems research, genetic algorithms and information systems. Syst. Res. Behav. Sci.: The Official J. Inter. Fed. Syst. Res. **17**, 149–162 (2000)

Chen, K.-Y., Wang, C.-H.: Support vector regression with genetic algorithms in forecasting tourism demand. Tour. Manage. **28**, 215–226 (2007)

Cheng, J.: Hybrid genetic algorithms for structural reliability analysis. Comput. Struct. **85**, 1524–1533 (2007)

Cheng, J.: An artificial neural network based genetic algorithm for estimating the reliability of long span suspension bridges. Finite Element. Anal. Design **46**, 658–667 (2010)

Chujai, P., Kerdprasop, N., Kerdprasop, K.:. Time series analysis of household electric consumption with ARIMA and ARMA models. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, 295–300 (2013)

Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J., Vapnik, V.: Support vector regression machines. Adv. Neural Info. Process. Syst., 155–161 (1997)

Ghosh, G., Mandal, P., Mondal, S.C.: Modeling and optimization of surface roughness in keyway milling using ANN, genetic algorithm, and particle swarm optimization. Int. J. Adv. Manufact. Tech. **100**(5–8), 1223–1242 (2017). https://doi.org/10.1007/s00170-017-1417-4

Hong, W.-C.: Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. Int. J. Elect. Pow. Energy Syst. **31**, 409–417 (2009)

Jiang, C.-S., Liang, G.-Q.: Modeling shear strength of medium-to ultra-high-strength concrete beams with stirrups using SVR and genetic algorithm. Soft. Comput. **25**, 10661–10675 (2021)

Karimi, H., Dastranj, J.: Artificial neural network-based genetic algorithm to predict natural gas consumption. Energy Syst. **5**(3), 571–581 (2014). https://doi.org/10.1007/s12667-014-0128-2

Kaytez, F., Taplamacioglu, M.C., Cam, E., Hardalac, F.: Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. Int. J. Elect. Pow. Energy Syst. **67**, 431–438 (2015)

Li, R., Chen, X., Balezentis, T., Streimikiene, D., Niu, Z.: Multi-step least squares support vector machine modeling approach for forecasting short-term electricity demand with application. Neural Comput. Appl. **33**(1), 301–320 (2020). https://doi.org/10.1007/s00521-020-04996-3

Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., Wei, Y.: A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Math. Comput. Model. **58**, 458–465 (2013)

Makridakis, S., Wheelwright, S., Hyndman, R.J.: Forecasting: Methods and Applications. Wiley (1998)

Nourani, V., Komasi, M., Mano, A.: A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water Resour. Manage **23**, 2877 (2009)

Nwankwo, S.C.: Autoregressive integrated moving average (ARIMA) model for exchange rate (Naira to Dollar). Acad. J. Interdis. Stud. **3**, 429 (2014)

Nyoni, T.: Modeling and Forecasting Naira/USD Exchange Rate in Nigeria: A Box-Jenkins ARIMA Approach (2018)

Onu, C.E., Nweke, C.N., Nwabanne, J.T.: Modeling of thermo-chemical pretreatment of yam peel substrate for biogas energy production: RSM, ANN, and ANFIS comparative approach. Appl. Surf. Sci. Adv. **11**, 100299 (2022)

Orjuela-Canon, A.D., Jutinico, A.L., Gonzalez, M.E.D., Garcia, C.E.A., Vergara, E., Palencia, M.A.: Time series forecasting for tuberculosis incidence employing neural network models. Heliyon 8 (2022)

Rapoo, M.I., Munapo, E., Chanza, M.M., Ewemooje, O.S.: Modelling and Forecasting Portfolio Inflows: A Comparative Study of Support Vector Regression, Artificial Neural Networks, and Structural VAR Models. Handbook of Research on Smart Technology Models for Business and Industry. IGI Global (2020)

Rapoo, M.I., Xaba, D.: A comparative analysis of artificial neural network and autoregressive integrated moving average model on modeling and forecasting exchange rate. Int. J. Econ. Manage. Eng. **11**, 2669–2672 (2017)

Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press (2014)

Smola, A.J., Scholkopf, B.: . Learning with Kernels. Citeseer (1998)

Suykens, J.A., van Gestel, T., de Brabanter, J., de Moor, B., Vandewalle, J.P.: Least Squares Support Vector Machines. World Scientific (2002)

Talei, A., Chua, L.H.C., Wong, T.S.: Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall–runoff modeling. J. Hydrol. **391**, 248–262 (2010)

Tay, F.E., Cao, L.: Application of support vector machines in financial time series forecasting. Omega **29**, 309–317 (2001)

Vapnik, V.: The Nature of Statistical Learning Theory. Springer Science & Business Media (2013)

Walls, L., Bendell, A.: Time series methods in reliability. Reliab. Eng. **18**, 239–265 (1987)
Xu, L., et al.: Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm. Energy **222**, 119955 (2021)
Zhiqiang, G., Huaiqing, W., Quan, L.: Financial time series forecasting using LPP and SVM optimized by PSO. Soft. Comput. **17**, 805–818 (2013)
Zhu, B., Ye, S., Wang, P., Chevallier, J., Wei, Y.M.: Forecasting carbon price using a multi-objective least squares support vector machine with mixture kernels. J. Forecast. **41**, 100–117 (2022)

Simulation for Analyzing Effect of Silver Meal Lot Sizing Rules in Bullwhip Effect

Liong Irena and I Gede Agus Widyadana(✉)

Industrial Engineering Department, Petra Christian University, Surabaya, Indonesia
gede@petra.ac.id

Abstract. Bullwhip effect is a variability of demand distorting phenomenon in supply chain. The difference in variability of demand is caused by varied methods in interpreting the data. One of the causes is the lot sizing rules which are used by the vendors. This research analyzes the effect of lot sizing rules on bullwhip effect on determine the order quantity with Silver Meal method. The supply chain model consists of two suppliers who are three retailers and one wholesaler. The bullwhip effect can be seen on the variability of wholesaler's demand. Bullwhip analysis is conducted using a simulation method. This paper will also show the impact of higher variability and the impact of holding cost. The result showed that Silver Meal method will produced higher variability on wholesaler demand. Using Silver meal on higher variability demand will decrease the bullwhip effect by only 1%. By reducing the holding cost it will increase the bullwhip effect, however the total cost will be lower.

Keywords: Simulation · Bullwhip effect · Lot sizing · Silver meal · Dynamic · Supply chain

1 Introduction

Bullwhip effect is distorted information that occurs in the supply chain. The distorted information are quantity order and quantity demand from a downstream to an upstream channel in the supply chain. The longer of the supply chain may cause the bullwhip effect will be even greater. Data received by the retailer will be different from the data received by wholesaler as well as distributors and factory. The order quantity is determined by considering the amount to be ordered and the inventory cost. The quantity to be ordered by the company is also consider the quantity of demand and the quantity for the inventory. The different methods used in determining the order quantity will produce different results for each supplier.

One of the most used methods is Silver Meal. This research is done by developing research of Pujawan [1] who researched the effect of lot sizing rules on one level supplier, and it is not known the impact of the bullwhip effect occurs at two levels, retailer, and wholesaler. Potter and Disney [2] derived closed form expression when demand is deterministic and the found that careful selection of batch size can impact bullwhip effect. Pujawan and Silver [3] developed heuristic procedures to select appropriate augmented

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 36–45, 2024.
https://doi.org/10.1007/978-3-031-50151-7_4

quantity for single item, discrete time, lot sizing situation. Wang and Disney [4] stated that ordering policies is one of element that build bullwhip effect. Relation between order batching to bullwhip effect was investigated and found that the relationship between batch size and demand amplification is non-monotonic [5].

In this paper we introduce simulation for analyzing the effect of lot sizing for bullwhip effects. Simulation has been used for analyzing bullwhip effect such as an application in hard goods retailer [5]. Silver meal lot sizing method is used since it one of lot sizing method that is used widely theoretically dan practically. Some theoretical analysis of Silver Meal is conducted by Govindan [6] and the application for single item dynamic lot sizing problem [7]. In this paper, the effect of Silver Meal methods to bullwhip effect is analyzed using three retailers and one wholesaler. This situation is different than Pujawan [1] that only considers one retailer and one wholesaler. A situation where there is more than one retailer is more applicable than only using one retailer. Research on the effect of lot sizing methods to bullwhip effect still has many attentions such as analyses in stochastic dynamic production [8], closed loop supply chain [9] and competing for supply and demand [10].

2 Problem Description

In this paper, the impact of the bullwhip effect that occurs in a supply chain caused by the lot sizing rules at two levels of supply chain is conducted. It is not known the impacts on each supplier if the buyer uses a set of different methods, inventory cost, and order variability. The purpose of this research is to know the impact of the bullwhip effect caused by the method of determining the quantity order by Silver Meal method in each retailer. Lee [11] stated that bullwhip effect is a phenomenon where the order variability received by the supplier will tend to greater from the seller to the buyer.

Order batching system is a way the company in making order. This order batching shows the ordering frequency and quantity. The ordering policy in a company can be done every day, every week, or every month. A company will order by considering the ordering cost and transportation cost. Most company do not have a certain ordering policy so if a variability of demand is low the company will do ordering with large amount.

Silver Meal is a method with an approach by finding a minimum average cost in each period. This method will produce a lot that can minimize the total costs per period. Demand from successive periods will accumulate into a lot size. Pujawan [1] examined the value of expected quantity that is the expected value of the quantity order at each period.

The expected quantity (*E(Q))* for Silver Meal can be expressed as follows:

$$E(Q) = (M - p)\mu \tag{1}$$

E(Q): Expected quantity.
M: Ordering period.
p: The probability of shortage.
μ: Mean of demand.

The order quantity is also affected by demand uncertainty and has to consider the variance of the order quantity. According Pujawan [1], the variance of the order quantity for Silver Meal can be obtain as:

$$Var(Q) = s^2 + p(1-p)\mu^2 - \frac{2s\mu}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\xi}{s}^2\right)\right) \tag{2}$$

Var(Q): The variance of order quantity.
s: standard deviation of demand.
p: probability of shortage.
μ: mean of demand
ξ: safety stock
The expected quantity and the variance of order quantity will be used to determine the order quantity. So that the order quantity can be obtained by:

$$\begin{aligned} Qt &= E(Q) + Z \times \sqrt{Var(Q) - I_{t-1}} \quad if\ D_t - I_{t-1} > 0 \\ Qt &= 0 \quad otherwise; \end{aligned} \tag{3}$$

Qt: Quantity order.
E(Q): Expected Quantity.
Z: Service level.
Var(Q): The variance of order quantity.
I_{t-1}: Inventory level at the end of period *t*.

3 Model Dan Result

3.1 Model Solution

The model settlement is done with two levels of retailer and wholesalers. There are three retailers and 1 wholesaler as shown in Fig. 1. All retailers will have different means and variance of demand. Then all the retailers will determine their quantity of orders depend on their demand in each period. All the order quantity by retailers in each period will be summed and become data of demand to wholesaler. The mean and variance of three retailers' successive sequence is 100-10; 200-20, and 500-50.

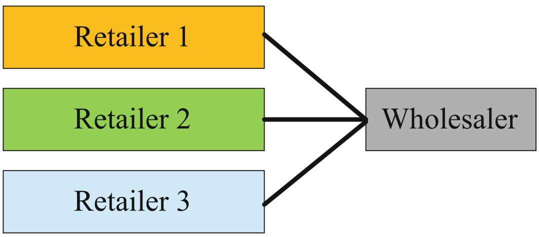


Fig. 1. Flow of supply chain

In this model the order quantity is counted in each period and calculate the cost per unit in each supplier. The calculation is simulated for 156 weeks and we assumption the ordering cost is \$400 and the holding cost is \$1 per unit.

Simulation step is as follows:

- Generating the demand of 3 retailers by using Random Number Generate based on mean and variance of demand in each retailer.
- Determine the quantity order in every period for each retailer. To determine the quantity order by using Silver Meal depend on the number of ordering period (M) for each supplier. The number of ordering period in each supplier will be different depend on the optimal cost per period. The number of order quantity can be determined by using Eq. (3).
- Calculating the number of inventories in each period by using equation as follows:

$$It = Numberofquantityorder_{(t)} - Numberofdemand_{(t)} + I_{(t-1)} \tag{4}$$

- Calculating the ordering cost and holding cost. There will be charged an ordering cost if there is an order in that period. So, the total of ordering cost for 156 weeks was obtained from the number of orderings is multiplied by the ordering cost. And holding cost is charged if there is unused inventory item in that period. Where is the total of holding cost was obtained from the number of inventories multiplied by the holding cost per unit.
- Calculate the demand of wholesaler which is obtained by summing the quantity order of three retailers in each period.
- Calculating the mean and standard deviation of demand for wholesaler that will be used to determine the order quantity. The same method as supplier is used to determine the order quantity at wholesaler.
- Calculate the ordering cost, holding cost, and the total cost per unit for wholesaler with the same method as retailers.

The first step to completing the model by using Silver Meal method by determining the number of ordering period (M). Where the number of ordering period (M) is based on comparison of cost per period for each additional period covered. Number of M is determined by calculating the cost per unit of the number of ordering period for $M =$ 2, 3, 4, 5 and so on until there is increasing in the cost per unit. The number of service levels that are used is 1. The result of determining the number of M with Silver Meal can be seen in Table 1.

Table 1. Result of the determination the number of M by Silver Meal

$H = 1$				
	Retailer 1	retailer 2	retailer 3	Wholesaler
M = 2	2.773	1.775	1.185	1.457
M = 3	2.642	1.970	1.574	1.955
M = 4	2.852	2.258	1.996	2.347

Table 1 shows the average cost per unit for each supplier at each number of M. The calculation is done with M equal to 2, 3, dan 4. The results show that for every additional period, so that the cost per unit will be higher with M = 3 except for retailer 1. Then for the retailer 1, we add the period with number of M equal to 4. The cost per unit for retailer 1 with $M = 4$ will be higher so the calculation is stopped. The results above show that retailer 1 is better to make an order with Silver Meal method with number of ordering period (M) is 3. While for retailer 2, retailer 3, and wholesaler are better to make an order with number of ordering period $M = 2$. This calculation is run 10 times to get the average cost per unit as shown in Table 2.

Table 2. Average of cost per unit ($/unit)

H = 1	Retailer 1	Retailer 2	Retailer 3	Wholesaler	Total cost
Silver meal	2.651	1.793	1.179	1.441	7.065

Table 3. Mean and standard deviation of demand retailer and wholesaler

	Retailer 1	Retailer 2	Retailer 3	Wholesaler
Mean	100	200	500	805.1
Standard deviation	10	20	50	598.77

Table 3 shows the results of the calculation of cost per unit from retailer 1 to retailer 3 show the change in cost per unit becomes lower. The cost per unit at the wholesaler will be higher than retailer 3 because of the higher variance of demand wholesaler then the quantity order in each period is unstable and often has shortage or overstocking.

Verification is a step to verify the model which has been made. The verification is conducted to test whether the models are in compliance or not. Verification is done by increasing and decreasing the standard deviation of demand. Small standard deviation is given by 5% of the mean of demand while large standard deviation given by 30% of the mean of demand as shown in Table 4.

Table 4. The number of changes standard deviation (unit).

	Mean	Stdev small	Stdev	Stdev large
Retailer 1	100	5	10	30
Retailer 2	200	10	20	60
Retailer 3	500	25	50	150

Table 5 shows that higher standard deviation results in higher cost. The result of this varication indicates that the models that has been verified.

Table 5. Verification (unit)

	Retailer 1	Retailer 2	Retailer 3
Small stdev	2,614	1,788	1,176
Initial stdev	2,651	1,793	1,179
Big stdev	2,793	1,952	1,318

The ordering cost and the holding cost to be one of the factors to be considered in making an order. The cost surely will affect the frequency and order quantity. The different comparison between the ordering cost and holding cost will affect the different bullwhip effect. In previous calculations is using \$400 for ordering cost and \$1 for holding cost. This simulation is done by changing the holding cost from \$1/unit to \$0,1/unit. And it affects the number of ordering period. The calculation result for determining the number of ordering period with the holding cost \$0,1/unit can be seen in Table 6.

Table 6. Result of determination the number of M with $H = \$0.1$/unit

	Retailer 1	Retailer 2	Retailer 3	Wholesaler
$M = 2$	2.100	1.089	0.485	0.596
$M = 3$	1.484	0.809	0.399	**0.341**
$M = 4$	1.466	0.799	**0.386**	0.359
$M = 5$	1.048	0.636	0.396	0,383
$M = 6$	0.949	0.625	0,412	0,464
$M = 7$	0.897	**0.621**	0,452	0,415
$M = 8$	0.875	0.626	0,484	0,454
$M = 9$	0.874	0,650	0,521	0,531
$M = 10$	**0.870**	0,690	0,569	0,486
$M = 11$	0.899	0,706	0,606	0,618

Table 7. Comparison of the cost per unit (\$/unit)

$H = \$1$				
Retailer 1	Retailer 2	Retailer 3	Wholesaler	Total cost
2.651	1.793	1.179	1.441	7.065
$H = \$0.1$				
0.886	0.618	0.384	0.655	2.543

Table 7 shows the average of cost per unit will be lower if the holding cost be reduced to \$0,1/unit. That is because the holding costs are cheaper so that each supplier is better at keeping inventory rather that ordering. This is because the number of ordering period is larger so that will save the ordering cost. This can be evidenced from the frequency of orders made for 156 weeks as seen in Table 8.

Table 8. Frequency of ordering (times)

	Retailer 1	Retailer 2	Retailer 3	Wholesaler
$H = 1$	79.0	80.8	79.8	78.0
$H = 0{,}1$	23.0	23.3	23.3	23.3

These results suggest that if the holding cost decrease so the frequency of ordering made will also be reduced, which saves about 50 times ordering. This will save the ordering cost because the holding cost is cheaper.

Sensitivity analysis is an analysis performed to determine the result of a change in a parameter. The sensitivity analysis performed in this research is the change of the cost of storage. This analysis to know the impact of changes on the holding cost ($0,1, $0,2, $0,5, $0,7, $0,8, and $1). This analysis is done to see whether changes in holding costs will have a significant impact on the cost per unit at each supplier.

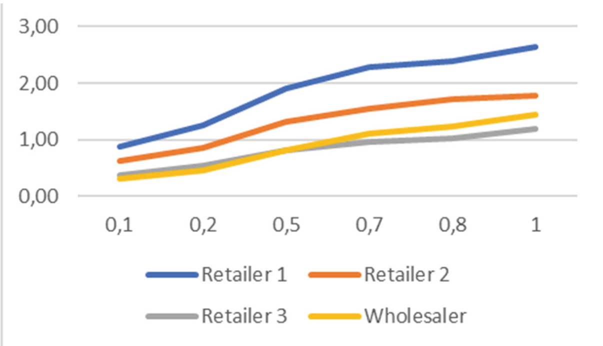


Fig. 2. Sensitivity analysis of holding cost ($/unit)

Figure 2 shows that the higher holding cost will produce cost per unit higher and any increase of holding cost will affect the cost significantly. This result also shows that with Silver Meal method cost will change constantly. These results also indicate that the changes of holding cost at wholesaler is smaller than retailer.

Bullwhip effect is caused by using Silver meal method where the ordering is done in a lot size (M) so the ordering not performed at each period. These is causing the demand at wholesaler is unstable on each period. Another thing that affects the bullwhip effect on wholesalers is the ordering of retailers are independent. There will be different impact of bullwhip effect if the wholesaler could know the real demand that needed by retailers and the wholesaler will combine the demands which are called aggregate planning. By aggregate planning, the wholesaler can adjust the requirement of the three retailers. This is also will affect the changes of mean and standard deviation of demand at wholesaler.

Table 9 shows the result of changes in the mean and standard deviation of demand wholesaler by aggregating and non-aggregating the demand. The mean and standard deviation of aggregate demand is smaller than non-aggregate demand. This is because the wholesaler combines the ordering of three retailers so that mean of demand at wholesaler will be smaller. By aggregating the demand, the standard deviation of demand will be smaller.

Table 9. The changes of mean and standard deviation of demand wholesaler (unit)

Mean of demand wholesaler	Mean aggregate	Non aggregate
Silver meal	276.7	805.1
Standard deviation of demand wholesaler	Mean aggregate	Non aggregate
Silver meal	556.81	598.77

Table 10. The cost per unit of wholesaler ($/unit)

	Mean aggregate	Non aggregate
Silver Meal	1,461	1,441

This also will impact on the cost per unit at wholesaler. Which the cost per unit at wholesaler with aggregate demand and non-aggregate will not different significantly (Table 10). So, by aggregating the demand, the mean of demand will smaller by 4 times of the non aggregate while the standard deviation will smaller by 10% of the non aggregate and the cost at wholesaler will not differ significantly.

The next simulation is to perform a calculation analysis if the variation of demand at each retailer be enlarged. The calculation of this simulation conducted to see the effect and changes that occur because of changes in variation of demand. The changes variation of demand is made by changing the standard deviation of demand at three retailers becomes larger, while the standard deviation at wholesaler will depend on quantity order of the three retailers as seen in Table 11.

Table 11. Changes standard deviation of retailers (unit)

	Mean	Initial stdev	Big stdev
Retailer 1	100	10	50
Retailer 2	200	20	100
Retailer 3	500	50	250

The changes of standard deviation of demand based on 50% of mean of demand at each retailer. Change in the standard deviation of retailer will certainly have an impact on the standard deviation of demand received at wholesaler. With other word, the bullwhip effect that received by wholesaler. The impact of bullwhip effect received by wholesaler can be seen on the standard deviation at wholesaler.

Table 12. Standard deviation of wholesaler (unit)

H = 1	Stdev	Large stdev
Silver meal	598.8	605.8
H = 0.1	Stdev	Large stdev
Silver meal	1071.5	1093.9

Table 12 shows the changes in standard deviation of demand at wholesaler that caused by enlarged the variation at retailers. With Silver Meal method, the standard deviation of demand at wholesaler will be even greater with an enlarged variation. However the impact of bullwhip effect is only increased ± 1% of the beginning variation. This is affecting the changes in cost per unit at each supplier.

Table 13. Changes of cost per unit at each suppliers ($/unit)

	$H = 1$		$H = 0{,}1$	
	Initial stdev	Large stdev	Initial stdev	Large stdev
Retailer 1	2.819	2.882	0.911	0.915
Retailer 2	1.783	1.968	0.619	0.634
Retailer 3	1.181	1.367	0.445	0.456
Wholesaler	1.458	1.300	0.443	0.438

Table 13 shows that with increasing the variation of demand at retailers there is no difference on ordering cost. While for the holding cost $1 then average changes of cost will only affect 10% of beginning cost.

4 Conclusion

Bullwhip effect is a phenomenon where there is a difference between the variation of demand at each channel within a supply chain. One of the factors that cause this difference is the method to determine the order quantity that are used by each supplier. This research analysis is conducted by developing two levels supply chain. The results show that Silver Meal method will cause the bullwhip effect is greater than the retailer to wholesaler. Where the standard deviation of demand received at wholesaler will be higher so that the ordering is done by wholesaler to the factory will also higher. By aggregating the demand of three retailers, the wholesaler will receive smaller mean and standard deviation but the cost per unit will be the same.

The impact of the bullwhip effect will also be influenced by the proportion of costs used. Where the greater gap between holding cost and ordering cost will affect the bullwhip effect produced by Silver Meal method, but the impact of the bullwhip effect generated by Silver Meal on a large variation will not increase much and only increase by 1% of beginning variation.

For future research, this research can be developed by considering more than two echelon player such as retailer, wholesaler dan factory with more than one player for each echelon, therefore the model is closer to reality.

References

1. Pujawan, I.N.: The effect of lot sizing rules on order variability. Eur. J. Oper. Res. **159**(3), 617–635 (2003)
2. Potter, A., Disney, S.M.: Bullwhip and batching: an exploration. Int. J. Prod. Econ. **104**(2), 408–418 (2006)
3. Pujawan, N., Silver, E.A.: Augmenting the lot sizing order quantity. Eur. J. Oper. Res. **188**(3), 705–722 (2008)
4. Wang, X., Disney, S.M.: The bullwhip effect: Progress, trends and directions. Eur. J. Oper. Res. **250**(3), 691–701 (2016)
5. Hussain, M., dan Drake, P.R.: Analysis of the bullwhip effect with order batching in multiple-echelon supply chain. Int. J. Phys. Dist. Logis. Manag. **41**(8), 797–814 (2011)
6. Nguyen, D.T., Adulyasak, Y., Landry, S.: Research manuscript: the bullwhip effect rule-based supply chain planning systems—a case-based simulation at a hard goods retailer. Omega **98**, 102121 (2021)
7. Govindan, K.: The optimal replenishment policy for time varying stochastic demand under vendor managed inventory **242**(2), 402–423 (2015)
8. Schulz, T.: A new Silver-Meal based heuristic for single-item dynamic lot sizing problem with returns and remanufacturing. Int. J. Prod. Res. **49**(9), 2519–2533 (2011)
9. Su, C.-T., Wong, J.-T.: Design of a replenishment system for a stochastic dynamic production/forecast lot-sizing problem under bullwhip effect. Expert Syst. Appl. **34**(1), 173–180 (2008)
10. Ponte, B., Dominguez, R., Cannella, S., dan Farminan, J.M.: The implications of batching in the bullwhip effect and customer service if closed-loop supply chain. Int. J. Prod. Econ. **244**, 108379 (2022)
11. Villa, S.: Competing for supply and demand: Understanding retailers' ordering decision. Int. J. Prod. Econ. **144**, 108355 (2022)

On Discontinuous Systems with Sliding Modes

Alexander Fominyh(✉)

Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, Saint Petersburg State University, Saint Petersburg, Russia
alexfomster@mail.ru

Abstract. The paper considers ODE systems with controls of a special form. In general case the controls are discontinuous in phase variables functions, so the system motion is going on in a so called sliding mode. It is known that such a motion may be described in terms of differential inclusions. A “principal" scheme for solving this differential inclusion is given. It is based on reducing the initial problem to minimizing a specially constructed functional. A detailed analysis of this functional differential properties is carried out, including the case when this functional appears to be nondifferentiable (but only superdifferentiable). An illustrative example is given.

Keywords: Sliding mode · Discontinuous system · Superdifferential

1 Introduction

In many practical problems, when trying to construct mathematical models of real physical processes, discontinuous in phase coordinates control is used; therefore, the right-hand sides of the systems of differential equations describing the process under study are discontinuous functions of the state vector.

In such systems the state vector may stay on one of the discontinuity surfaces (or on some their intersection) over a period of time of nonzero measure. Motion along the discontinuity surfaces or along some their intersection is called a sliding mode. Therefore, a corresponding definition of a system solution in case when the sliding mode occurs was introduced (see [1, 2]). Some of numerous practical applications of the sliding modes are presented in [3–5].

This paper aims at the problem of finding a solution of a differential inclusion describing the system motion in a sliding mode. Herewith, we will additionally impose some restrictions on the desired trajectory, for example, hitting a certain point at the final moment of time (when moving along the discontinuity surface). This inclusion has a special structure and in some cases some nonsmooth (in the phase coordinates) functions can appear in its right-hand side. Contributions of this paper are as follows. The superdifferentiability of the functional constructed is proved. A method for both finding a trajectory moving in a sliding mode and a sliding mode design problem is developed.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 46–56, 2024.
https://doi.org/10.1007/978-3-031-50151-7_5

2 Basic Definitions and Notations

In the paper we will use the following notations. $C_n[0, T]$ denotes the space of n-dimensional continuous on $[0, T]$ vector-functions. $P_n[0, T]$ is the space of piecewise continuous and bounded on $[0, T]$ n-dimensional vector-functions. In the paper we will also require the space $L_n^2[0, T]$ of square-summable in $[0, T]$ n-dimensional vector-functions. Let X be a normed space, then $|| \cdot ||_X$ denotes the norm in this space and X^* denotes the space conjugate to the space X.

In the paper we assume that each trajectory $x(t)$ is a piecewise continuously differentiable vector-function with bounded derivative in its domain. With the assumptions and the notations made we can suppose that the vector-function $x(t)$ belongs to the space $C_n[0, T]$ and that the vector-function $\dot{x}(t)$ belongs to the space $P_n[0, T]$.

For the arbitrary set $F \subset R^n$ let us define a support function of the vector $\psi \in R^n$ as $c(F, \psi) = \sup_{f \in F} \langle f, \psi \rangle$ where $\langle a, b \rangle$ is a scalar product of the vectors $a, b \in R^n$. Let also S_n be a unit sphere in R^n with the center in the origin. Let the vectors $\mathbf{e_i}$, $i = \overline{1, n}$, form the standard basis in R^n. 0_n denotes a zero element of a functional space of some n-dimensional vector-functions, and $\mathbf{0_n}$ — a zero element of the space R^n. Let E_m denote an identity matrix and $\mathbf{O_m}$ — a zero matrix in the space $R^m \times R^m$, let also diag[P, Q] denote a diagonal matrix with the blocks P and Q (where P and Q are the matrices of some dimensions). Denote $|x| = \sum_{i=1}^n |x_i|$, where $x \in R^n$.

In the paper we will use both superdifferentials of functions in a finite-dimensional space and superdifferentials of functionals in a functional space.

Consider the space $R^n \times R^n$ with the standard norm. Let $d = [d_1, d_2] \in \in R^n \times R^n$ be an arbitrary vector. Suppose that at the point (x, z) there exists such a convex compact set $\overline{\partial} h(x, z) \subset R^n \times R^n$ that

$$\frac{\partial h(x, z)}{\partial d} = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \big(h(x + \alpha d_1, z + \alpha d_2) - h(x, z) \big) = \min_{w \in \overline{\partial} h(x,z)} \langle w, d \rangle. \tag{1}$$

In this case the function $h(x, z)$ is called superdifferentiable at the point (x, z), and the set $\overline{\partial} h(x, z)$ is called a superdifferential of the function $h(x, z)$ at the point (x, z).

Consider the space $C_n[0, T]$ with the $L_n^2[0, T]$ norm. Let $g \in C_n[0, T]$ be an arbitrary vector-function. Suppose that at the point x there exists a convex weakly* compact set $\overline{\partial} I(x) \subset \big(C_n[0, T], || \cdot ||_{L_n^2[0,T]} \big)^*$ such that

$$\frac{\partial I(x)}{\partial g} = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \big(I(x + \alpha g) - I(x) \big) = \min_{w \in \overline{\partial} I(x)} w(g). \tag{2}$$

In this case the functional $I(x)$ is called superdifferentiable at the point x, and the set $\overline{\partial} I(x)$ is called a superdifferential of the functional $I(x)$ at the point x.

3 Statement of the Problem

Consider the system of differential equations

$$\dot{x} = Ax + Bu \tag{3}$$

with the initial point

$$x(0) = x_0 \tag{4}$$

and with the desired endpoint

$$x_j(T) = x_{Tj}, \quad j \in J. \tag{5}$$

In formula (3) A is a constant $n \times n$ matrix, B is a constant $n \times m$ matrix. For simplicity suppose that $B = \mathrm{diag}[E_m, \mathbf{O_{n-m}}]$. The system is considered on the given finite time interval $[-t^*, T]$ (here T is a given final moment; see comments on the moment t^* below). Then by Sect. 2 assumption $x(t)$ is an n-dimensional continuous vector-function of phase coordinates with a piecewise-continuous and bounded on $[-t^*, T]$ derivative; the structure of the m-dimensional control u will be specified below. In formula (4) $x_0 \in R^n$ is a given vector; in formula (5) x_{Tj} are given numbers, corresponding to those coordinates of the state vector which are fixed at the right endpoint, here J is a given index set.

Let also the discontinuity surface

$$s(x, c) = \mathbf{0_m}, \quad s(x, c) = (s_1(x, c), \dots, s_m(x, c))', \tag{6}$$

be given, $s(x, c)$ is a continuously differentiable vector-function. We assume that the structure of the surface is given while the parameters $c \in R^\ell$ are unknown.

Let us immediately write out the explicit form of the control which is mainly used in this paper. Let

$$u_i = -\alpha_i |x| \mathrm{sign}(s_i(x, c)), \quad i = \overline{1, m}, \tag{7}$$

where $\alpha_i \in [\underline{a}_i, \overline{a}_i]$, $i = \overline{1, m}$, are some positive finite numbers (with $\underline{a}_i$, $\overline{a}_i$, $i = \overline{1, m}$, given) which are sometimes called gain factors.

In book [6] it is shown that if surface (6) is a hyperplane, then under natural assumptions and with sufficiently large values of the factors α_i, $i = \overline{1, m}$, controls (7) ensure system (3) hitting a small vicinity of discontinuity surface (6) from arbitrary initial state (4) in the finite time t^* and further staying in this neighborhood with the fulfillment of the condition $s_i(x(t)) \to 0$, $i = \overline{1, m}$, at $t \to \infty$. We assume that all the conditions required are already met, i. e. the numbers $\underline{a}_i$, $\overline{a}_i$, $i = \overline{1, m}$, are sufficiently large. In Sects. 4, 5, 6 we will be interested in the behavior of the system on the discontinuity surface (on the time interval $[0, T]$).

On the surfaces $s_i(x, c) = 0$, $i = \overline{1, m}$, the solution of the original discontinuous system is a solution of the following differential inclusion:

$$\dot{x}_i \in A_i x + [\underline{a}_i, \overline{a}_i]|x|[-1, 1] = A_i x + [-\overline{a}_i, \overline{a}_i]|x|, \quad i = \overline{1, m}. \tag{8}$$

$$\dot{x}_i = A_i x, \quad i = \overline{m+1, n}. \tag{9}$$

In formulas (8), (9), A_i is the i-th row of the matrix A, $i = \overline{1, n}$. Note that the more detailed version of the definition of a discontinuous system solution is given in [2]. It has a strict and rather complicated form so we don't consider the details here. For our purposes it is sufficient to postulate that the system (3) solutions with controls (7) are the solutions of system (8), (9) by definition.

Let $F(x) = (f_1(x), \ldots, f_n(x))'$, where $f_1(x), \ldots, f_n(x)$ run through the corresponding sets $F_1(x), \ldots F_n(x)$ from the right-hand sides of inclusions (8); herewith $f_i(x) = F_i(x) := A_i x$, $i = \overline{m+1, n}$ (see (9)), so we can write

$$\dot{x} \in F(x). \tag{10}$$

We formulate the problem as follows: it is required to find such a trajectory $x^* \in C_n[0, T]$ (with the derivative $\dot{x}^* \in P_n[0, T]$) which moves along discontinuity surface (6) (in practice — in some sufficiently small vicinity of this surface) while $t \in [0, T]$ (the parameters $c^* \in R^\ell$ are to be determined as well), satisfies differential inclusion (10) and boundary conditions (4), (5). We suppose that there exists such a solution. A justification (from the engineering point of view) of the solution existence may be found in [6].

Remark 1 Instead of trajectories from the space $C_n[0, T]$ with derivatives from the space $P_n[0, T]$ one may consider absolutely continuous on the interval $[0, T]$ trajectories with measurable and almost everywhere bounded on $[0, T]$ derivatives, what is more natural for differential inclusions. The choice of the solution space in the paper is explained by the possibility of its practical construction.

4 Reduction to a Variational Problem

Insofar as $\forall x \in R^n$ the set $F(x)$ is a convex compact set in R^n, then inclusion (10) may be rewritten as follows:

$$\dot{x}_i(t)\psi_i \le c(F_i(x(t)), \psi_i) \quad \forall \psi_i \in S_1, \quad \forall t \in [0, T], \quad i = \overline{1, n}.$$

The support function of the set $F_i(x)$ can be expressed by the formula

$$c(F_i(x), \psi_i) = \psi_i A_i x + \overline{a}_i |x||\psi_i|, \quad i = \overline{1, m},$$

$$c(F_i(x), \psi_i) = \psi_i A_i x, \quad i = \overline{m+1, n}.$$

We see that the support function of the set $F_i(x)$ is continuously differentiable in the phase coordinates x if $x_i \neq 0$, $i = \overline{1, n}$.

Denote $z(t) = \dot{x}(t)$, $z \in P_n[0, T]$, then from (4) one has

$$x(t) = x_0 + \int_0^t z(\tau) d\tau.$$

Put

$$\ell_i(\psi_i, x, z) = \langle z_i, \psi_i \rangle - c(F_i(x), \psi_i),$$

$$h_i(x, z) = \max_{\psi_i \in S_1} \max\{0, \ell_i(\psi_i, x, z)\}, \quad h(x, z) = (h_1(x, z), \ldots, h_n(x, z))'$$

and construct the functional

$$\varphi(z) = \frac{1}{2}\int_0^T h^2\Big(x_0 + \int_0^t z(\tau)d\tau, z(t)\Big)dt. \tag{11}$$

It is not difficult to check that for functional (11) the relation

$$\begin{cases} \varphi(z) = 0, \text{ if } \dot{x}_i(t)\psi_i \le c(F_i(x(t)), \psi_i) \quad \forall \psi_i \in S_1, \quad \forall t \in [0, T], \quad i = \overline{1, n}. \\ \varphi(z) > 0, \text{ otherwise,} \end{cases}$$

holds true, i. e. inclusion (10) takes place iff $\varphi(z) = 0$.

Introduce the functional

$$\chi(z) = \frac{1}{2}\sum_{j\in J}\Big(x_{0j} + \int_0^T z_j(t)dt - x_{Tj}\Big)^2. \tag{12}$$

It is seen that condition (4) on the left endpoint is automatically satisfied due to the vector-function $z(t)$ definition and condition (5) on the right endpoint is satisfied iff $\chi(z) = 0$.

As noted above, the search for a solution of a differential inclusion is carried out on the discontinuity surfaces $s_i(x, c) = 0$, $i = \overline{1, m}$, so also introduce the functional

$$\omega(z, c) = \frac{1}{2}\int_0^T s^2\Big(x_0 + \int_0^t z(\tau)d\tau, c\Big)dt. \tag{13}$$

Construct the functional

$$I(z, c) = \varphi(z) + \chi(z) + \omega(z, c). \tag{14}$$

So the original problem has been reduced to minimizing functional (14) on the space $P_n[0, T] \times R^\ell$. Denote z^*, c^* a global minimizer of this functional. Then $x^*(t) = x_0 + \int_0^t z^*(\tau)d\tau$ is a solution of the initial problem (and the vector c^* defines the discontinuity surface structure).

Remark 2 The structure of the functional $\varphi(z)$ is natural as the value $h_i(x(t), z(t))$, $i = \overline{1, n}$, at each fixed $t \in [0, T]$ is just the Euclidean distance from the point $z_i(t)$ to the set $F_i(x(t))$; functional (11) is half the sum of squares of the deviations in $L_n^2[0, T]$ norm of the trajectories $z_i(t)$ from the sets $F_i(x)$, $i = \overline{1, n}$, respectively; the meaning of functionals (12), (13) structures is obvious.

5 Necessary Minimum Conditions of the Functional *I*(*z*, *c*) in a Particular Case

It is obvious that the point x^* is the original problem solution iff the functional $I(z, c)$ vanishes at the corresponding point, i. e. $I(z^*, c^*) = 0$. In order to obtain a more constructive minimum condition (useful to develop a numerical method for solving the problem), the differential properties of the functional $I(z, c)$ are to be studied. Then under the assumption made, it may be proved that the functional $I(z, c)$ is Gâteaux differentiable. Denote ψ_i^* such a value that $\max\{0, \ell_i(\psi_i^*(x, z), x, z)\} = \max_{\psi_i \in S_1} \max\{0, \ell_i(\psi_i, x, z)\}$, $i = \overline{1, n}$ (and note [7] that $\psi_i^*(x, z)$ is unique in the case $h_i(x, z) > 0$). The following theorem formulates the known minimum condition for a Gateaux differentiable functional.

Theorem 1 Let the trajectories $x_i(t)$, $i = \overline{1, n}$, vanish only at isolated time moments of the interval $[0, T]$. In order for the point (z^*, c^*) to minimize the functional $I(z, c)$, it is necessary that

$$\begin{aligned} 0_n \times \mathbf{0}_\ell = \Bigg[&\sum_{i=1}^{n} h_i(x^*, z^*)\psi_i^*(x^*, z^*)\mathbf{e_i} \\ &- \sum_{i=1}^{n} \int_t^T h_i(x^*(\tau), z^*(\tau)) \frac{\partial c_i(F_i(x^*(\tau)), \psi_i^*(x^*(\tau), z^*(\tau)))}{\partial x} d\tau \\ &+ \sum_{j \in J} \left(x_{0j} + \int_0^T z_j^*(t)dt - x_{Tj} \right) \mathbf{e_j} \\ &+ \sum_{i=1}^{m} \int_t^T s_i(x^*(\tau), c^*) \frac{\partial s_i(x^*(\tau), c^*)}{\partial x} d\tau, \\ &\sum_{i=1}^{m} \int_0^T s_i(x^*(\tau), c^*) \frac{\partial s_i(x^*(\tau), c^*)}{\partial c} d\tau \Bigg] \end{aligned}$$

where 0_n is a zero element of the space $P_n[0, T]$.

Proof The detailed proof is carried out similarly as in [7] (see also [8, 9]).

Remark 3 Note that the trajectory functional space is not closed in $L_n^2[0, T]$ metric so formally some optimization method can lead to unacceptable points at some iterations. But since this functional space is everywhere dense in the space $L_n^2[0, T]$ in practice we approximate such a point with an acceptable one.

Example 1 Consider the following system

$$\dot{x}_1(t) \in [-1, 1](|x_1(t)| + |x_2(t)| + |x_3(t)|), \quad \dot{x}_2(t) = x_1(t) - x_2(t), \quad \dot{x}_3(t) = x_2(t)$$

with the boundary conditions

$$x_1(0) = -3, \ x_2(0) = 4, \ x_3(0) = 6, \quad x_1(1) = 0.$$

Let the discontinuity surface in this example be of the form

$$s(x, c) = x_1 + c_1 x_2 + c_2 = 0.$$

A simplest steepest descent method (in the functional space) [10] was used in order to minimize the functional $I(z, c)$ in this problem.

The point $(z_{\{0\}}, c_{\{0\}}) = (0, 0, 0, 1.25, -1.25)'$ was taken as the initial one (here $(0, 0, 0)'$ is a zero point in the space $P_3[0, 1]$ and the point $(1.25, -1.25)'$ belongs to the space R^2). Herewith, we have $I(z_{\{0\}}, c_{\{0\}}) \approx 37.28125$.

At the 60-th iteration the point $(z_{\{60\}}, c_{\{60\}})'$ was constructed and we approximately put $(z^*, c^*) = (z_{\{60\}}, c_{\{60\}})$. Herewith $c^* \approx (0.98467, -0.93868)'$, and we have $I(z^*, c^*) \approx 0.00015$.

Having obtained the parameters c^*, put $x_1(t) := -c_1^* x_2(t) - c_2^*$ and substitute them into the last two equations of the system given. Integrate this closed-loop system with the Runge-Kutta 4–5-th order method, then check that the first inclusion is also satisfied and finally have $x_1(1) \approx -0.00431$, so we see that the desired value on the right endpoint is achieved with an error of the order 5×10^{-3}. So this value has been improved via "correcting" the parameters of the surface considered. Picture 1 demonstrates the results of calculations. The black lines denote the curves which were obtained via the method of the paper, while the dashed red lines denote the curves obtained via integration of the closed-loop system (and via the relation $x_1(t) = -c_1^* x_2(t) - c_2^*$).

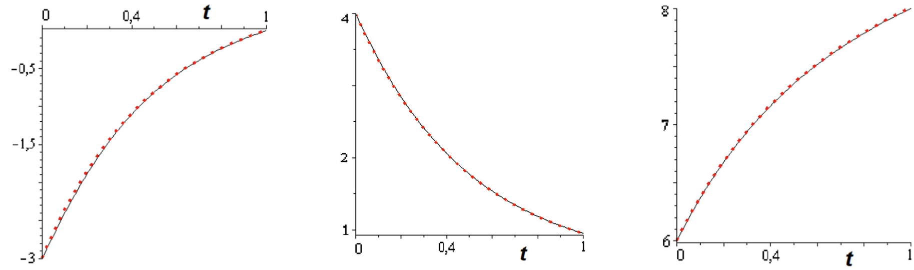


Fig. 1. Example 1. The trajectories $x_1^*(t)$, $x_2^*(t)$, $x_3^*(t)$.

6 Differential Properties of the Functional $I(x, z, c)$ in a More General Case

In this section we do not to "go to the space of derivatives" but consider x and z as independent variables and work with the following functional (we keep the previous notation for it):

$$I(x, z, c) = \varphi(x, z) + \chi(x) + \omega(x, c),$$

where as previously (we omit the $1/2-$factor in the functional $\varphi(x)$ for its more convenient exploration below)

$$\varphi(x, z) = \int_0^T h^2\big(x(t), z(t)\big)dt, \tag{15}$$

$$\chi(z) = \frac{1}{2} \sum_{j \in J} \left(x_{0j} + \int_0^T z_j(t)dt - x_{Tj} \right)^2,$$

$$\omega(x, c) = \frac{1}{2} \int_0^T s^2(x(t), c)dt.$$

In Sect. 5 it was shown that the functional $I(z, c)$ is Gâteaux differentiable, and according to the assumption made there, the case was excluded when at least one of the phase variables would vanish on some interval of the set $[0, T]$ of nonzero length. Now assume that some of the trajectories can be identically equal to zero on some nonzero time interval of the segment $[0, T]$ and study the differential properties of the functional $I(x, z, c)$ in this case.

For simplicity consider the case $n = 2$ and only the functions $\ell_1(\psi_1, x, z)$ and $h_1(x, z)$ (here we denote them $\ell(\psi_1, x_1, x_2, z_1)$ and $h(x_1, x_2, z_1)$ respectively) and the time interval $[t_1, t_2] \subset [0, T]$ of nonzero length; the general case is considered in a similar way. Then we have $\ell(\psi_1, x_1, x_2, z_1) = z_1\psi_1 - a_1x_1\psi_1 - a_2x_2\psi_1 - -b|x_1||\psi_1| - b|x_2||\psi_1|$, where $a_1 := a_{1,1}$, $a_2 := a_{1,2}$, $b := \overline{a}_1$. Fix some point $(x_1, x_2) \in R^2$. Let $x_1(t) = 0$, $z_1(t) = 0$, $x_2(t) > 0$ at $t \in [t_1, t_2]$; other cases are studied in a completely analogous fashion.

(a) Suppose that $h_1(x, z) > 0$, i. e. $h_1(x, z) = \max_{\psi_1 \in S_1} \ell_1(\psi_1, x, z) > 0$.

Our aim is to apply the corresponding theorem on a directional differentiability from [11]. The theorem of this book considers the inf-functions so we will apply this theorem to the function $-\ell(\psi_1, x_1, x_2, z_1)$. For this check that the function $h(x_1, x_2, z_1)$ satisfies the following conditions:

(i) the function $\ell(\psi_1, x_1, x_2, z_1)$ is continuous on $S_1 \times R^2 \times R$;
(ii) there exist a number β and a compact set $\mathcal{C} \in R$ such that for every (x_1, x_2, z_1) in the vicinity of the point $(0, x_2, 0)$ the level set

$$lev_\beta(-\ell(\cdot, x_1, x_2, z_1)) = \{\psi_1 \in S_1 \mid -\ell(\psi_1, x_1, x_2, z_1) \le \beta\}$$

is nonempty and is contained in the set $\mathcal{C}$;
(iii) for any fixed $\psi_1 \in S_1$ the function $\ell(\psi_1, \cdot, \cdot, \cdot)$ is directionally differentiable at the point $(0, x_2, 0)$;
(iv) if $d = [d_1, d_2] \in R^2 \times R$, $\gamma_n \downarrow 0$ and ψ_{1_n} is a sequence in $\mathcal{C}$, then ψ_{1_n} has a limit point $\overline{\psi}_1$ such that

$$\limsup_{n\to\infty} \frac{-\ell(\psi_{1_n}, 0 + \gamma_n d_{1,1}, x_2 + \gamma_n d_{1,2}, 0 + \gamma_n d_2) - (-\ell(\psi_{1_n}, 0, x_2, 0))}{\gamma_n} \ge \frac{\partial(-\ell(\overline{\psi}_1, 0, x_2, 0))}{\partial d},$$

where $\dfrac{\partial \ell(\overline{\psi}_1, 0, x_2, 0)}{\partial d}$ is the derivative of the function $\ell(\overline{\psi}_1, x_1, x_2, z_1)$ at the point $(0, x_2, 0)$ in the direction d.

The verification of conditions (i), (ii) is obvious.

In order to verify condition (iii), it suffices to note that since $b > 0$, then at the fixed $\psi_1 \in S_1$ the function $-b|x_1||\psi_1|$ is superdifferentiable (and hence is differentiable in directions) at the point 0, herewith, its superdifferential at the point 0 is the segment $\mathrm{co}\{-b|\psi_1|, b|\psi_1|\}$. An explicit expression for the derivative of this function at the point 0 in the direction $d_{1,1}$ is $-b|d_{1,1}||\psi_1|$.

Finally, check condition (iv). In fact, the equality will be proved there. Let $[d_1, d_2] \in R^2 \times R$, $\gamma_n \downarrow 0$ and ψ_{1_n} is some sequence from $\mathcal{C}$. Calculate

$$\limsup_{n\to\infty} \frac{\ell(\psi_{1_n}, 0 + \gamma_n d_{1,1}, x_2 + \gamma_n d_{1,2}, 0 + \gamma_n d_2) - \ell(\psi_{1_n}, 0, x_2, 0)}{\gamma_n}$$
$$= \limsup_{n\to\infty}$$
$$\frac{\gamma_n d_2 \psi_{1_n} - \gamma_n a_1 d_{1,1} \psi_{1_n} - \gamma_n a_2 d_{1,2} \psi_{1_n} - b|\gamma_n d_{1,1}||\psi_{1_n}| - b\gamma_n d_{1,2}|\psi_{1_n}|}{\gamma_n}$$
$$= \limsup_{n\to\infty} \left(d_2 \psi_{1_n} - a_1 d_{1,1} \psi_{1_n} - a_2 d_{1,2} \psi_{1_n} - b|d_{1,1}||\psi_{1_n}| - b d_{1,2} |\psi_{1_n}|\right).$$

Let $\overline{\psi}_1$ be a limit point of the sequence ψ_{1_n}. Then by the directional derivative definition we have

$$\frac{\partial \ell(\overline{\psi}_1, 0, x_2, 0)}{\partial d} = d_2 \overline{\psi}_1 - a_1 d_{1,1} \overline{\psi}_1 - a_2 d_{1,2} \overline{\psi}_1 - b|d_{1,1}||\overline{\psi}_1| - b d_{1,2} |\overline{\psi}_1|.$$

From last two equalities one obtains that condition iv) is fulfilled.

Thus, the function $h(x_1, x_2, z_1)$ satisfies conditions i)-iv), so it is differentiable in directions at the point $(0, x_2, 0)$ [11], and its derivative in the direction d at this point is expressed by the formula

$$\frac{\partial h(0, x_2, 0)}{\partial d} = \sup_{\psi_1 \in \mathcal{S}(0, x_2, 0)} \frac{\partial \ell(\psi_1, 0, x_2, 0)}{\partial d}$$

where $\mathcal{S}(0, x_2, 0) = \arg\max_{\psi_1 \in S_1} \ell(\psi_1, 0, x_2, 0)$. However, as shown above, in the considered problem the set $\mathcal{S}(0, x_2, 0)$ consists of the only element $\psi_1^*(0, x_2, 0)$, hence

$$\frac{\partial h(0, x_2, 0)}{\partial d} = \frac{\partial \ell(\psi_1^*(0, x_2, 0), 0, x_2, 0)}{\partial d}.$$

Finally, recall that by the directional derivative definition one has the equality

$$\frac{\partial \ell(\psi_1^*(0, x_2, 0), 0, x_2, 0)}{\partial d} = d_2 \psi_1^* - a_1 d_{1,1} \psi_1^* - a_2 d_{1,2} \psi_1^* - b|d_{1,1}||\psi_1^*| - b d_{1,2} |\psi_1^*|$$

where we have put $\psi_1^* := \psi_1^*(0, x_2, 0)$.

From last two equalities we finally obtain that the function $h(x_1, x_2, z_1)$ is superdifferentiable at the point $(0, x_2, 0)$, but it is also positive in this case, hence the function $h^2(x_1, x_2, z_1)$ is superdifferentiable at the point $(0, x_2, 0)$ as a square of a superdifferentiable positive function (see [12]).

(b) In the case $h_1(x, z) = 0$ it is obvious that the function $h^2(x_1, x_2, z_1)$ is differentiable at the point $(0, x_2, 0)$ and its gradient vanishes at this point.

Remark 4 The proof above may be significantly simplified if one uses the fact that $\psi_1, \psi_{1_n}, \overline{\psi}_1, \psi_1^* \in S_1$. However, since the statement about the superdifferentiability of functions having an analogous structure is of independent interest, so such a way of proof is used which may be applied to the more general case when ψ_1 belongs to an arbitrary compact subset of the space R (therefore, we consciously did not put $|\psi_1| = |\psi_{1_n}| = |\overline{\psi}_1| = |\psi_1^*| = 1$ in the proof).

Above we have considered a particular case. Dividing the interval $[0, T]$ into the segments in which some of the phase trajectories vanish and some retain a definite sign (we suppose that there is a finite number of such segments) and arguing in each of these intervals similarly and also using superdifferential calculus rules [12], write down the superdifferential of the function $h_i(x, z)$ (in the case $h_i(x, z) > 0$). One has

$$\overline{\partial} h_i(x, z) = \psi_i^* \mathbf{e_{i+n}} - \psi_i^* [A_i', \mathbf{0_n}] + \sum_{j=1}^{n} \overline{\partial}(-\overline{a}_i |x_j| |\psi_i^*|), \quad i = \overline{1, m}$$

$$\overline{\partial} h_i(x, z) = \psi_i^* \mathbf{e_{i+n}} - \psi_i^* [A_i', \mathbf{0_n}], \quad i = \overline{m+1, n}.$$

So in this case the integrand of functional (15) is superdifferentiable as a square of a superdifferentiable nonnegative function [12]. It turns out that this fact implies that functional (15) itself is superdifferentiable in this case.

Theorem 2 Let the interval $[0, T]$ may be divided into a finite number of intervals, in every of which each phase trajectory is either identically equal to zero or retains a certain sign. Then the functional $\varphi(x, z)$ is superdifferentiable, i.e.

$$\frac{\partial \varphi(x, z)}{\partial g} = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \big(\varphi((x, z) + \alpha g) - \varphi(x, z)\big) = \min_{w \in \overline{\partial} \varphi(x,z)} w(g). \tag{16}$$

Here the set $\overline{\partial} \varphi(x)$ is of the following form

$$\overline{\partial} \varphi(x, z) = \Bigg\{ w \in \Big(C_n[0, T] \times P_n[0, T], || \cdot ||_{L_2^2 n[0,T]}\Big)^* \ \Big| $$
$$w(g) = \int_0^T \langle w_1(t), g_1(t) \rangle dt + \int_0^T \langle w_2(t), g_2(t) \rangle dt \quad \forall g \in C_n[0, T],\ g_2 \in P_n[0, T],$$
$$w_1(t), w_2(t) \in L_n^\infty[0, T], \quad [w_1(t), w_2(t)] \in \overline{\partial} h^2(x(t), z(t)) \quad \forall\, t \in [0, T] \Bigg\}. \tag{17}$$

Proof Let us give just the scheme of proof for brevity:

(1) With the use of the superdifferential definition (see formula (1)), Lebesgue's dominated convergence theorem (considering special structure of the function $h^2(x, \dot{x})$ and continuity of the function $\psi^*(x, \dot{x})$) and Filippov lemma, we first show that the direction derivative of the functional $\varphi(x)$ is of form (16).

(2) Denote $X = \Big(C_n[0, T] \times P_n[0, T], || \cdot ||_{L_n^2[0,T]}\Big)$. Now our aim is to show that the set $\overline{\partial} \varphi(x, z)$ in formula (1) is convex and weakly* compact in the space X^* (see (2)):
  (a) convexity follows from convexity of the function $h^2(x, \dot{x})$ superdifferential;
  (b) we show that the set $\overline{\partial} \varphi(x, z)$ is bounded in $L_n^2[0, T]$-norm via superdifferentiability of the function $h^2(x, \dot{x})$ and upper-semicontinuity of the mapping $t \to \overline{\partial} h^2(x(t), \dot{x}(t))$;
  (c) recall the following fact: if w_n is the sequence of functions from the set $\overline{\partial} \varphi(x)$ converging to the function w^* in the strong topology of the space $L_n^2[0, T]$, then this sequence has the subsequence w_{n_k} converging pointwise to w^* almost everywhere on $[0, T]$; using this fact and the function $h^2(x, \dot{x})$ superdifferential definition we show that the set $\overline{\partial} \varphi(x, z)$ is weakly closed in the space $L_n^2[0, T]$;

(d) finally, we use the isometric isomorphism of the spaces X^* and $L_n^2[0, T]$, the fact that here weak compactness is equivalent to weak* compactness in the space $L_n^2[0, T]$ and the fact that weak compactness is equivalent to boundedness in norm and weak closedness in the space $L_n^2[0, T]$ in order to finish the proof.

7 Conclusion and Acknowledgments

Thus, in the paper: (1) a discontinuous system was studied as a differential inclusion, (2) finding a trajectory moving in a sliding mode and a sliding mode design was reduced to a variational problem (3) the functional obtained differential properties were studied, (4) a numerical method for the problem was developed.

The main results of this paper (Sects. 5, 6) were obtained in IPME RAS and supported by Russian Science Foundation (project no. 20-71-10032).

References

1. Filippov, A.F.: Differential Equations with Discontinuous Right-Hand Side. Nauka, Moscow (1985)
2. Aizerman, M.A., Pyatniskii, E.S.: Fundamentals of the theory of discontinuous systems. I. Automa. Remote Control. **7**, 33–47 (1974)
3. Ashrafiuon, H., Muske, K.R., McNinch, L.C., Soltan, R.A.: Sliding-mode tracking control of surface vessels. IEEE Trans. Ind. Electron. **55**(11), 4004–4012 (2008)
4. Beltran, B., Ahmed-Ali, T., Benbouzid, M.: High-order sliding-mode control of variable-speed wind turbines. J. IEEE Trans. Ind. Electron. **56**(9), 3314–3321 (2009)
5. Tang, Y.: Terminal sliding mode control for rigid robots. Automatica **34**(1), 51–56 (1998)
6. Utkin, V.I.: Sliding Modes in Control and Optimization. Nauka, Moscow (1981)
7. Fominyh, A.V.: Method for finding a solution to a linear nonstationary interval ODE system. Vestn. St. Petersburg Univ. Appl. Math. Comput. Sci. Control Process. **17**(2), 148–165 (2021)
8. Fominyh, A.V.: A numerical method for finding the optimal solution of a differential inclusion. Vestnik St. Petersburg Univ., Math. **51**(4), 397–406 (2018)
9. Fominyh, A.V.: A method for solving differential inclusions with fixed right end. Vestnik St. Petersburg Univ. Appl. Math. Comput. Sci. Control Process. **14**(4), 302–315 (2018)
10. Kantorovich, L.V., Akilov, G.: Functional Analysis. Nauka, Moscow (1977)
11. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Science+Business Media Publ, New York (2000)
12. Demyanov, V.F., Rubinov, A.M.: Basics of Nonsmooth Analysis and Quasidifferential Calculus. Nauka, Moscow (1990)

Carbon Footprint Optimization for Efficient Effluent Treatment Selection by Using Fuzzy Optimization Modelling

Kenneth Tiong Kim Yeoh[1](✉), Chee Ming Choo[1], Mohd Faizan Jamaluddin[1], Geen Yue Lim[1], Jun Hao Chew[1], and Lik Yin Ng[2]

[1] Centre for Water Research, Faculty of Engineering, Built Environment and Information Technology, SEGi University, 47810 Kota Damansara, Selangor, Darul Ehsan, Malaysia
kennethyeoh@segi.edu.my

[2] Department of Chemical & Petroleum Engineering, Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, Malaysia

Abstract. With rising industrial ESG needs, companies must optimize their effluent treatment plant design to reduce carbon emissions and meet government regulations. Industrial decision-makers face challenges in selecting the best effluent treatment technology that meets all goals and regulations. By considering the energy usage of various technological combinations, a model is created to minimize the carbon footprint. Fuzzy optimization modelling and carbon footprint integration result in more effective and sustainable industrial effluent treatment systems. In the future, the multi-objective optimization will be considered to balance environmental sustainability and practicality.

Keywords: Effluent treatment plant · Carbon footprint · Fuzzy optimization modeling · Optimization

1 Introduction

Companies would optimize their effluent treatment plant design to reduce carbon emissions and meet government regulations. Industrial efforts follow governmental effluent discharge regulations and increasing limits on environmental sustainability [1]. Due to the wide range of options available, industrial decision-makers face challenges in selecting the best effluent treatment technology that meets these objectives and regulations. An effective design must be planned before installation to avoid costly corrections later. The combination of various technologies in an effluent treatment plant (ETP) affects its energy consumption and carbon footprint. A model can optimize the carbon footprint by considering power consumption leading towards carbon footprint calculations. The term "carbon footprint" refers to equivalent carbon dioxide (eCO_2) and greenhouse gasses (GHG) emissions. The lowering of eCO_2 emissions has made sustainable production a priority in the effluent treatment industry.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 57–63, 2024.
https://doi.org/10.1007/978-3-031-50151-7_6

2 Methodology

2.1 Framework

Optimizing effluent treatment plant, ETP designs involve analyzing literature and generic data as outlined in a decision-making flowchart framework.

1. Collect data for effluent treatment case study.
2. Develop a generic ETP superstructure based on design or literature reviews.
3. Conduct material mass balance, calculate eCO_2 and conduct fuzzy optimization analysis.
4. Code equations into LINGO and solve the optimization model.
5. Select optimised pathway(s) that fulfilled decision constraints (Fig. 1).

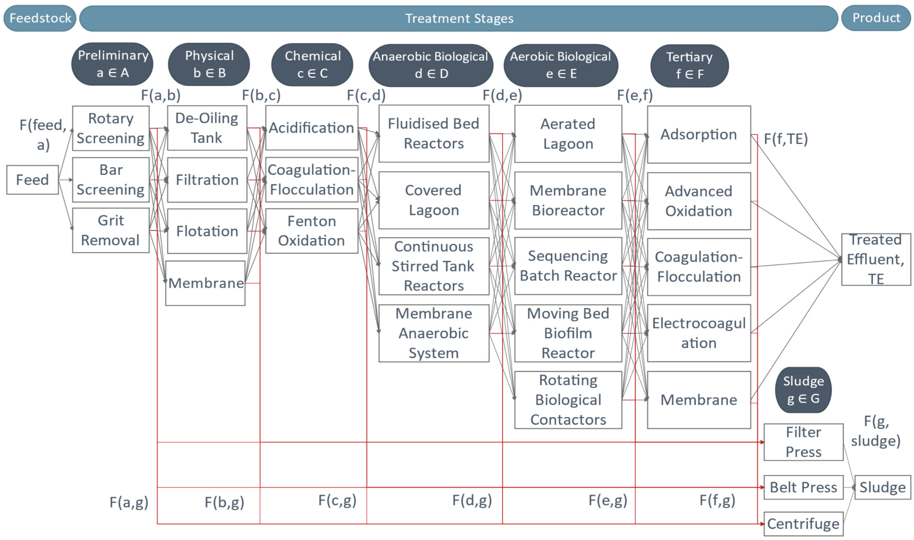


Fig. 1. ETP superstructure developed by incorporating each treatment stage as outlined.

2.2 Multi-objective Mathematical Model

The methodology involves programming using LINGO to optimize the selection of ETP pathways based on linear programming. The design model considers flowrate, material mass balance, and carbon footprint based on [2, 3].

Material Mass Balance. Material mass balance depends on volumetric flowrates and component loading valances. For each repeating treatment stage j, flowrates in litres per day between preceding and current ($F_{i,j}$), current and subsequent ($F_{j,k}$) stages, as shown in Fig. 2.

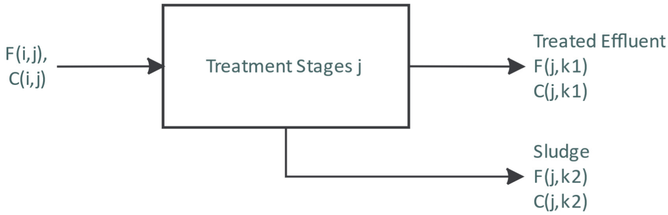


Fig. 2. Individual treatment stage input-output balance.

In Eq. (1), the flowrate factor, $P_j^{flowrate}$, represents a fraction of treated effluent flowrate to inlet flowrate. The model has zero degrees of freedom. The flowrate balance Eq. (2) applies for each treatment stage j. The treated effluent flowrate from one stage equals the next stage's inlet flowrate. C_m represents the stream concentration of contaminations such as chemical oxygen demand (COD), biological oxygen demand (BOD), and total suspended solids (TSS) in kg/L [3]. FC_m is the stream component loading, while treatment stage j removal efficiency of component m, $n_{m,j}$, is defined by Eq. (3) [4] and is obtained from the literature. The component loading in the sludge can be determined using $n_{m,j}$ with Eq. (4) [2]. Equation (5) is the balance equation for the component m in each treatment stage j.

$$F_{j,k1} = P_j^{flowrate} F_{i,j}, \forall j \tag{1}$$

$$F_{i,j} = P_j^{flowrate} F_{i,j} + F_{j,k2}, \forall j \tag{2}$$

$$n_{m,j} = 1 - C_{j,k1} \div C_{i,j}, \forall m, j \tag{3}$$

$$F_{j,k2} C_{m,j,k2} = F_{i,j} C_{m,i,j} n_{m,j}, \forall m, j \tag{4}$$

$$F_{i,j} C_{m,i,j} = P_j^{flowrate} F_{i,j} C_{m,j,k1} + F_{i,j} C_{m,i,j} n_{m,j}, \forall m, j \tag{5}$$

Equation (6) is the ETP overall volumetric flowrate balance. This study considers the flowrates across the superstructure with a, b, c, d, e, f ∈ J and g ∈ k2. Although tertiary treatment does not generate sludge, it does contribute to material mass balance. The treated effluent concentration, $C_{m,f,TE}$, must meet the effluent concentration standard, $C_{m,std}$, as Eq. (7) [3] to ensure that the chosen ETP system pathway produces a qualified treated effluent.

$$F_{feed,a} = F_{f,TE} + \sum_{j=1}^{J} F_{j,k2}, \forall j \tag{6}$$

$$C_{m,f,TE} \leq C_{m,std}, \forall m \tag{7}$$

Carbon Footprint. In this study, the carbon footprint is calculated indirectly from the power consumption of each treatment stage. Equation (8) is used to calculate stage j, the carbon footprint (kg eCO_2/d) of stage j using the inlet flowrate of the treatment stage (L/d) and the power consumption (kWh/L) of each technology which depends on the carbon footprint factor, P^{carbon} (kg eCO_2/kWh) that measure the amount of carbon dioxide equivalent released per unit of power consumed. The total carbon footprint for the system, TotCarbon (kg eCO_2/d), is calculated by summing the carbon footprint of each technology as shown in Eq. (9) [3].

$$Carbon_j = F_{i,j} E_j^{power} P^{carbon}, \forall j \tag{8}$$

$$TotCarbon = \sum_{j=1}^{J} Carbon_j \tag{9}$$

Fuzzy Optimization Model

The fuzzy optimization model incorporates a Lambda coefficient, the degree of satisfaction, λ, which is a parameter with a value between 0 and 1. Higher λ represents higher satisfaction [3]. By allowing for uncertainty and imprecision in decision-making and approximating human reasoning processes, this coefficient enables the handling of imprecise or uncertain information and enhances adaptability and flexibility to real-world scenarios. Equation (10) was the objective function for the optimization problem in this study. The decision-makers can determine the limits of the total carbon footprint, $TotCarbon^{UL}$ and $TotCarbon^{LL}$, based on their interests. *UL* and *LL* denotes upper and lower limits. The function maximize λ in Eq. (11) to find the optimum pathway with the minimum carbon footprint based on selected degree of satisfaction in multi objective studies.

$$\frac{TotCarbon^{UL} - TotCarbon}{TotCarbon^{UL} - TotCarbon^{LL}} \geq \lambda \tag{10}$$

$$Maximize\, \lambda \tag{11}$$

Constraints for optimization. The relationships between the inlet and outlet flowrate and component m concentration are defined in Eqs. (12) and (13) for individual treatment stages. Equation (14) defines the relationship between the inlet and outlet component m loading of the overall ETP system. The conditions of these equations ensure that selected pathways flowrates and concentrations meet all requirements.

$$F_{i,j} > F_{j,k1} > F_{j,k2} > 0, \forall j \tag{12}$$

$$C_{m,j,k2} \geq C_{m,i,j} \geq C_{m,j,k1} \geq 0, \forall m, j \tag{13}$$

$$F_{feed,a} C_{m,feed,a} \geq F_{g,sludge} C_{m,g,sludge} > F_{f,TE} C_{m,f,TE} \tag{14}$$

3 Results and Discussion

The hypothetical example, ETP to treat 150000 L/d of highly polluted oleochemical effluent (40000 mg/L COD, 20000 mg/L BOD, 15000 mg/L TSS) to comply with Malaysia's Department of Environment regulations (200 mg/L COD, 50 mg/L BOD, 100 mg/L TSS) [1, 5]. Table 1 indicates data on the technology's removal efficiency, power consumption, and each treatment stage's flowrate factor. This study uses a carbon footprint factor of 0.433 kg eCO2/kWh and sets carbon emission cap allowances of 10000 kg eCO_2/year and 4,000 kg eCO_2/year (upper and lower limits, respectively).

The case study was optimized using LINGO coding. An optimum treatment pathway with a minimized carbon footprint was obtained with fuzzy optimization optimization (Fig. 3).

This case study, with a 0.76 of degree of satisfaction, demonstrates the potential of the fuzzy optimization model in multi-objective studies. The fuzzy optimization model was successful in reducing the carbon footprint of an effluent treatment plant by 83% (or 12.4 kg eCO_2/d) compared to the worst scenario and by 16% (or 2.4 kg eCO_2/d) compared to the average scenario, resulting in a total of 882.4 kg eCO_2 carbon equivalent reduction each year.

4 Conclusion

Integrating fuzzy optimization modelling and optimization with a carbon footprint perspective in industrial effluent treatment has proven to be effective in ensuring sustainable and responsible operations in line with ESG goals. Fuzzy optimization modelling provides a flexible and robust tool for modelling complex and uncertain systems, while the carbon footprint perspective brings in a critical sustainability aspect that considers the impact of the treatment processes for ESG. Integrating these two approaches leads to developing efficient treatment systems thssat effectively remove pollutants while minimizing the carbon footprint. This approach promotes the shift towards a green economy.

Table 1. Effluent treatment technology specifications [5].

Stage	Technology	n_{COD}	n_{BOD}	n_{TSS}	E^{power} (kWh/L)	Pflowrate
Preliminary, $a \in A$	Rotary Screening	0.3	0.285	0.36	0.0000070	0.999
	Bar Screening	0.28	0.266	0.336	0.0000084	0.999
	Grit Removal	0.3	0.285	0.36	0.0000063	0.999
Physical, $b \in B$	De-Oiling Tank	0.9	0.855	0.95	0.0000056	0.99
	Filtration	0.6	0.57	0.72	0.0000064	0.99
	Flotation	0.7	0.665	0.84	0.0000096	0.99
	Membrane	0.75	0.7125	0.9	0.0000072	0.99
Chemical, $c \in C$	Acidification	0.46	0.437	0.552	0.0001890	0.985
	Coagulation-Flocculation	0.66	0.627	0.792	0.0002268	0.985
	Fenton Oxidation	0.76	0.722	0.912	0.0001701	0.985
Anaerobic biological, $d \in D$	Fluidized Bed Reactors	0.5	0.475	0.6	0.0000319	0.98
	Covered Lagoon	0.68	0.646	0.816	0.0000348	0.98
	Continuous Stirred Tank Reactors	0.72	0.684	0.864	0.0000232	0.98
	Membrane Anaerobic System	0.8	0.76	0.96	0.0000261	0.98
Aerobic biological, $e \in E$	Aerated Lagoon	0.66	0.627	0.792	0.0000290	0.975
	Membrane Bioreactor	0.78	0.741	0.936	0.0000203	0.975
	Sequencing Batch Reactor	0.02	0.019	0.024	0.0000383	0.975
	Moving Bed Biofilm Reactor	0.04	0.038	0.048	0.0000255	0.975
	Rotating Biological Contactors	0.8	0.76	0.96	0.0000287	0.975
Tertiary, $f \in F$	Adsorption	0.7	0.665	0.84	0.0000074	0.97
	Advanced Oxidation	0.6	0.57	0.72	0.0000110	0.97
	Coagulation-Flocculation	0.66	0.627	0.792	0.0000083	0.97

(*continued*)

Table 1. (*continued*)

Stage	Technology	n_{COD}	n_{BOD}	n_{TSS}	E^{power} (kWh/L)	Pflowrate
	Electrocoagulation	0.75	0.7125	0.9	0.0000079	0.97
	Membrane	0.8	0.76	0.85	0.0000119	0.97
Sludge, g ∈ G	Filter Press	–	–	–	0.0000084	0.97
	Belt Press	–	–	–	0.0000096	0.97
	Centrifuge	–	–	–	0.0000108	0.97

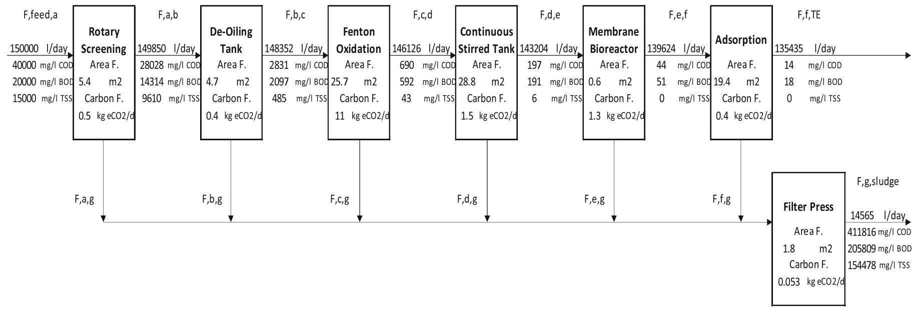


Fig. 3. The optimized effluent treatment design, chosen from 10800 potential pathways, consists of bar screening, de-oiling tank, coagulation-flocculation, covered lagoon, rotating biological contactors, membrane, and belt press, which produces a qualified effluent with COD at 13.8 mg/L, BOD at 17.5 mg/L, TSS at 0.1 mg/L, and 14.97 kg eCO_2/d emission.

References

1. Department of Environment: Environmental quality (sewage) regulations. Ministry of Environment and Water (PU(A) 432) (2009)
2. Ho, J.Y., Wan, Y.K., Andiappan, V., Ng, D.K.S.: Material flow cost account-based approach for synthesis and optimisation of wastewater treatment plant. Chem. Eng. Trans. **76**, 529–534 (2019). Available at: https://doi.org/10.3303/CET1976089
3. Ho, J.Y., Ooi, J., Wan, Y.K., Andiappan, V.: Synthesis of a sustainable effluent treatment plant for sago industry using fuzzy optimization. Chem. Eng. Trans. **83**, 373–378 (2021)
4. Von Sperling, M., Verbyla, M.E., Oliveira, S.M.: Assessment of Treatment Plant Performance and Water Quality Data: A Guide for Students, Researchers and Practitioners. IWA publishing (2020)
5. Department of Energy: Benchmark model for wastewater treatment using an activated sludge process (2022). Available at: https://catalog.data.gov/dataset/benchmark-model-for-wastewater-treatment-using-an-activated-sludge-process

Application of Analytical Hierarchy Process (AHP) in Assessing the Risk of COVID-19 Contraction by the Urban Public Through Transport Services

Trust Tawanda[1], Godfrey Muzuka[2], and Elias Munapo[3](✉)

[1] Department of Statistics and Operations Research, National University of Science and Technology, Bulawayo, Zimbabwe
[2] Institute of Development Studies, National University of Science and Technology, Bulawayo, Zimbabwe
[3] School of Economics and Decision Sciences, North West University, Mafikeng Campus, Mafikeng, South Africa
elias.munapo@nwu.ac.za

Abstract. Transport has emerged as one the factors increasing the risk of contracting COVID-19 virus hence the study deliberated on determining the most safe transport people can use during the pandemic period. This was determined amongst five transport operators in the city of Bulawayo, Zimbabwe which include, Zimbabwe United Passenger Company (ZUPCO) Buses, ZUPCO Minibuses, ZUPCO Omnibuses and Non-ZUPCO operators. Data was collected through a combination of independent observations. Observations were based on administration of COVID-19 prevention measures thus Temperature screening (TS), Sanitizer Administration (SA), Social Distancing (SD), Disinfection (D) and Proper Mask Wearing (PMW). These were in five categories ZUPCO Buses (N = 3452), ZUPCO Omnibus (N = 4450), ZUPCO Minibuses (N = 1339), Non-ZUPCO Omnibuses (N = 3950) and Pirating small cars (N = 2300). Binary recording system was used to capture administration of COVID-19 prevention measures (1 for administration and 0 otherwise). Analytical Hierarchy Process (AHP) was used to rank the transport operators in terms of COVID-19 prevention measures criterions. The computed composite weights are as follows ZUPCO buses (0.169), ZUPCO Omnibuses (0.174), ZUPCO Minibuses (0.290), Non-ZUPCO Omnibuses (0.052) and Pirating small cars (0.310). It was observed that Pirating small cars are the safest transport to use, however very few private car owners were willing to offer the service during the peak of the pandemic. This means the need for considering other alternatives in the following order of safety ZUPCO Minibuses, ZUPCO Omnibuses, ZUPCO Buses and lastly Non-ZUPCO Omnibuses. The contributions of this study are as follows (i) transport users are informed on which transport category has least risk of contracting COVID-19 and (ii) policy makers and law enforcement agents are hinted on grey areas in relation to transport operators that are lagging behind in terms of implementation of COVID-19 prevention measures.

Keywords: Equality COVID-19 · Lockdown · Risk · Transport services · Analytical Hierarchy Process (AHP)

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 64–74, 2024.
https://doi.org/10.1007/978-3-031-50151-7_7

1 Introduction and Background

Urban centers globally are regarded as main hot spot areas when it comes to the COVID-19 pandemic. There are several reasons why this is the case and among them is that, they are the centers of commerce and connection points between national, regional and international cities. Another issue is that of higher mobility associated with urban residents due to work related reasons and in Zimbabwean context, high unemployment has seen urban residents engaging in street vending which necessitates their presence in the city center on daily basis. Unlike in the developed world regions where policies and resources allows for the transportation of people to worry less about their safety in relation to contraction of COVID-19, the case in developing nations like Zimbabwe is different. COVID-19 regulations have reduced transport operators and residents of large cities are found stranded to make their way to and from work. Trapped in this situation, most residents ended up valuing getting to the intended destination with limited consideration on how high is the risk of contracting this deadly disease through using a given mode of available transport than the other. Transportation of people is no doubt a major factor in the spreading of COVID-19 amongst urban residents. The situation is worse in Zimbabwe where the authorized Zimbabwe United Passengers Company (ZUPCO) doesn't have enough fleet to timeously ferry people to their intended destinations. Other operators opted out of the ZUPCO contract citing several reasons joining the so called Mushikashika (unauthorized) transport operators. The public faced a difficult decision to make on the transporter to use when factors to do with cost, time, general safety, and most importantly for this study the risk of contracting the COVID-19 pandemic is considered. The decision here should be made on the operator to use where the risk of contracting the pandemic is least out of the following options; Bus (75 seater), Mini-bus (35 seater), Commuter Omnibus (14-Seater), Pirating small cars. These options carry different COVID-19 contraction risk levels which are as a function of multiple criteria as capacity, ventilation, social distancing, sanitizer administration, temperature checking, disinfecting and mask wearing. This presents a Multi-Criteria Analysis (MCA) problem faced by the public in minimizing chances of them contracting the pandemic on daily basis as they travel to and from work.

2 Related Literature

The Analytical hierarch process (AHP) method was developed by Saaty in 1980. Mohammed and Daham [1] used AHP to evaluate flipped classrooms in Iraqi secondary schools. The research revealed that students and teachers in Iraqi secondary schools preferred flipped classroom learning compared to conventional cognitive learning. Briceno-Leon *et al.* [2] selected the best pumping stations using AHP. Technical, economic factors, investment costs, maintenance costs, and operational cost were considered in the selection process. Zekhnini *et al.* [3] applied AHP to analyze supply chain risks such as cybersecurity risk and information risk. The research is important in that it helps organizations to have secure chains. Dar *et al.* [4] calculated weights for thematic layers for ground water potential zones using AHP. Layers were classified as excellent, good, moderate and others. AHP was used to analyze Covid -19 pandemic impact on energy project performance by Hussain *et al.* [5]. It was determined that personal factors and government measures are the factors that influenced project performance the most. This study helps on project implementation and management during and after COVID-19 pandemic time. Saha *et al.* [6] applied AHP to assess the impact of COVID-19 on the tourism sector in 50 selected countries. Countries which include China, Malaysia, South Korea, Russia, and Portugal among the 50 considered countries. The study revealed that COVID-19 pandemic affected the economies of countries that depends on tourism sector. AHP has many important applications which include, head of warehouse selection considered by Dewi *et al.* [7], employee selection in Hermawan *et al.* [8], thesis session pass recommendation in Valentino *et al.* [9], land evaluation in Chaudhary *et al.* [10], Green economic efficiency assessment in Naseer *et al.* [11], Evaluation of barrier for sustainable e-learning in Naveed *et al.* [12], E-wallet payment evaluation in Fam *et al.* [13], Resilience assessment to public health disaster in Liu *et al.* [14], determining public opinion during distribution of COVID -19 vaccines in Klumpp *et al.* [15], determining priority factors during production and distribution of face mask in Kumar *et al.* [16], identifying regions vulnerable to COVID -19 pandemic transmission in Gupta *et al.* [17] among other applications. This shows how useful AHP is in dealing with multi-criteria decision making challenges.

3 Analytical Hierarch Process Methodology

AHP methodology requires that firstly, the problem is decomposed into a hierarchy of criterions and alternatives. Secondly, criterion and alternatives pairwise comparisons based on the fundamental scale of absolute numbers. Finally, alternatives are ranked subjected to criterions. The aim of this study is to assess the risk of COVID-19 contraction by the urban public through transport services. Figure 1 summaries the problem by decomposing it into a hierarchy (Table 1).

To make comparisons, we need a scale of numbers that indicate how many times more important or dominant one element is over another element with respect to the criterion or property with respect to which they are compared.

3.1 Consistency Index and Consistency Ratio

Consistency ratio (CR) is a measure how consistency the judgements have been relative to large samples of purely random judgements. Consistency ratio is a comparison between Consistency index (CI) and Random Consistency index (RI). Consistency index is measure of consistency as a deviation or degree of consistency. λ_{Max} is the maximum eigenvalue of the matrix and is needed to calculate Consistency index. $\lambda_{Max} - n$ measures deviation of the judgements from the consistent approximation. CI and CR are given by (1) and (2) below (Table 2).

$$CI = \lambda_{Max} - n \tag{1}$$

$$CR = \frac{CI}{RI} \tag{2}$$

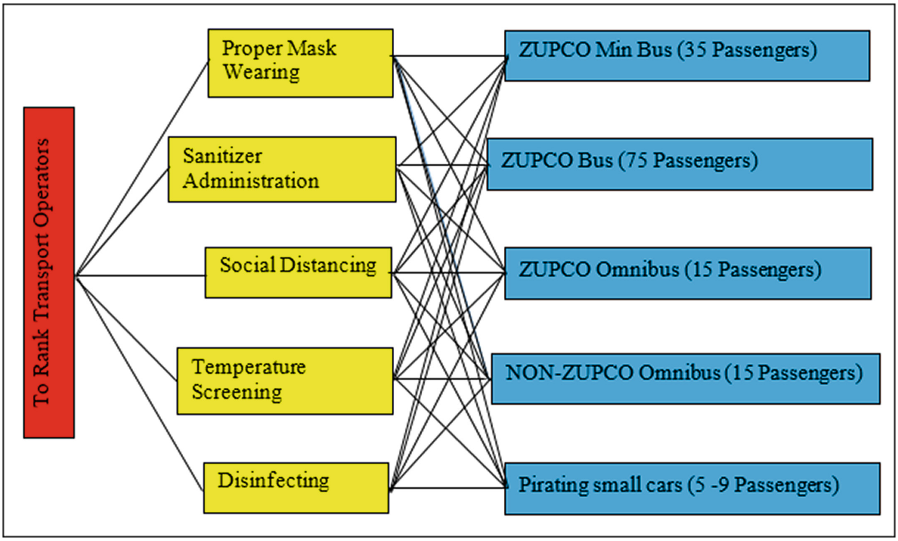


Fig. 1. Decomposition of a problem into a hierarchy

Table 1. The fundamental scale of absolute numbers

Numerical scale	Interpretation
1	Equally important
3	Weakly important
5	Moderately important
7	Very important
9	Extremely Important
2, 4, 6, 8	In between their two neighbors

Table 2. Mean random consistency index

n	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

3.2 Data Collection Procedure

The study was based on primary quantitative data collected through participant observations from January 2020 to August 2021. Data collection was done at 100% compliance with COVID-19 regulations especially that to do with reducing physical contact by distance of not less than 1 m (social distancing) and proper mask wearing. The need to observe these regulations necessitated scrapping off participation by the public (face to face interviews) as well as use of questionnaires as these were deemed unsafe in the process of administering and collection after completion. Face to face interviews would enable transport users to provide their opinions on which of the operator do they find most safe and least safe in relation to the risk of them contracting COVID-19, however the need for 100% compliance with the pandemic's regulations made these not feasible for the study to employ. Data collection was a continuous process from the onset of the pandemic through to the third wave/variant lockdown period. Participant observations by default limit much interaction with people and therefore they were the cornerstone for the data collection process. Participant observation is a process by which an assessment team collects data on naturally occurring behavior within their usual context.. Active participation was employed as it allowed gaining access to the setting and observing the group under study. It also allowed moving around at will observing in detail and depth and in different situations.

Both Probability and Non-probability sampling methods were used to select points across the city of Bulawayo where the observations were carried out. A total of twelve (12) observation points were purposively selected across the city. These included the main depot for Zimbabwe United Passengers Company, three undesignated picking and dropping off points and eight designated points. These points were opted for as they allowed for easy access to all sorts transporters around the city. Researchers randomly visited each of these sites during peak hour periods on working days observing administration of COVID-19 regulations. A checklist was used to record Category of operator under observation and the administration or non-administration of each of the regulations during the boarding and dropping off process. Transport operators were stratified first according to authorization to transport people into Authorized (ZUPCO operators) and Unauthorized (Non-ZUPCO operators). Further stratification was done based on capacity leading to sub classes as ZUPCO Bus (75 seater), ZUPCO Mini bus (35 seater), ZUPCO Omnibus (15 seater), Non-ZUPCO Omnibus (15 seater) and Pirating small cars (5–9 seater). Total figures for each category of operators (alternatives/options) under consideration were deemed not necessary but the number of independent observations made on administration of Covid-19 regulations /Criteria in the form of; Disinfecting of vehicles, Temperature screening of passengers, Proper mask wearing by operators and customers, Sanitizing, and Social distancing were of paramount importance in achieving the aim of the study. Independent observations made with respect to the five operators under study are shown in Table 3.

Table 3. Sample observations per operator

Category of operators/alternatives/options	Number of observations
ZUPCO buses	3452
ZUPCO omnibuses	4450
ZUPCO minibuses	1339
Non-ZUPCO omnibuses	3950
Pirating small cars	2300

Before covid-19 pandemic and the re-introduction of the Zimbabwe United Passengers Company (ZUPCO) in the transportation of urban residents to and from work and shopping among other reasons for going to town, commuter omnibuses, private cars and minibuses were the main operators. Buses came as an initiative by the government to phase out omnibuses and their number is still low in the business. Minibuses' number even before the pandemic was lesser than that of omnibuses and private cars. This influenced the frequency of observations made on these operators with omnibuses leading as they dominate in this business.

3.3 Description of Criteria and Data Capturing Procedure

Disinfecting

Disinfecting in this study referred to spraying of chemicals like Hydrogen Peroxide with the potential of killing the virus inside the vehicle or any other part that encounters passengers. Observations were made and recorded on whether operators disinfected vehicles every time before making a new trip at picking up points. Where this was done, a yes was put and where it was not done, a no was placed under the observed operator on a checklist. This was done for each criterion under every observed transport operator on every observation session.

Temperature screening

Describes a process of body temperature measurement by transport operators on each customer entering the vehicle to eliminate those whose body temperature would have exceeded 37 °C.This was observed and recorded for all operators taking note of whether operators administer it or not.

Sanitizing

The spraying or application of sanitizer (spirit liquids) in the hands of both operators and customers during the boarding process. Operators were observed to see if all people were sanitized or not in their hands before boarding.

Proper mask wearing

Observations were made to see if all operators and customers boarding were wearing masks covering the nose and mouth. Since COVID -19 is an airborne kind of infection, improper mask wearing by one person would risk others in contracting the virus. Where this situation existed, the criteria was concluded not fully observed and the opposite is true in cases where it was not observed.

Social distancing

Researchers observed whether operators and customers observed 1 m apart at picking up points and inside the vehicles or were not doing so.

4 Analysis

The data collected was summarized as percentages on the administration of COVID-19 regulations by different operators as shown in Table 4 on alternatives indicators. Pairwise comparisons (Tables 5, 6, 7, 8, 9, and 10) were made based on alternatives indicators.

Table 4. Alternative indicators

Alternatives/options	PMW	SA	SD	TS	D	Number of observations
ZUPCO buses	255 (7.39%)	2782 (80.59%)	55 (1.59%)	179 (5.19%)	1320 (38.24)	3452
ZUPCO omnibuses	221 (4.97%)	3319 (74.58%)	150 (3.37%)	325 (7.30%)	1801 (40.47%)	4450
ZUPCO minibuses	147 (10.98%)	961 (71.77%)	57 (4.26%)	117 (8.74%)	732 (54.67%)	1339
Non-ZUPCO omnibuses	86 (2.18%)	29 (0.73%)	39 (0.99%)	5 (0.13%)	13 (0.33%)	3950
Pirating small cars	1111 (48.30%)	166 (7.23%)	0 (0.00%)	3 (0.13%)	0 (0.00%)	2300

4.1 Pairwise Comparisons

Table 5. Criteria pairwise comparisons

Criteria pairwise comparison	PMW	SD	SA	TS	D	Priority	Rank
Proper mask wearing (PMW)	1	3	4	6	9	0.4908	1
Social distancing (SD)	1/3	1	3	4	6	0.2566	2
Sanitizer Administration (SA)	1/4	1/3	1	2	5	0.1352	3
Disinfecting (D)	1/6	1/4	1/2	1	3	0.0802	4
Temperature screening (TS)	1/9	1/6	1/5	1/3	1	0.037	5
Consistency Ratio = 0.06 < 10%, Acceptable							

5 Results and Discussions

Research findings (Summarized in Table 11) have shown that in relation to enforcement of proper mask wearing regulation, pirating small cars have the highest weight (0.58) whilst Non-ZUPCO Omnibuses have least weight (0.042). In terms of sanitizer administration, ZUPCO buses have the highest weight (0.432) yet pirating small cars have

Table 6. Alternatives pairwise comparisons with respect to proper mask wearing

Proper mask wearing (PMW)	ZB	ZO	ZM	NZO	PSC	Priority	Rank
ZUPCO buses (ZB)	1	2	1/3	3	1/7	0.1	3
ZUPCO omnibuses (ZO)	1/2	1	1/4	2	1/8	0.064	4
ZUPCO minibuses (ZM)	3	4	1	5	1/5	0.208	2
Non-ZUPCO omnibuses (NZO)	1/3	1/2	1/5	1	1/9	0.042	5
Pirating small cars (PSC)	7	8	5	9	1	0.58	1
Consistency Ratio = 0.08 < 10%, Acceptable							

Table 7. Alternatives pairwise comparisons with respect to sanitizer administration

Sanitizer administration (SA)	ZB	ZO	ZM	NZO	PSC	Priority	Rank
ZUPCO buses (ZB)	1	2	3	9	7	0.432	1
ZUPCO omnibuses (ZO)	1/2	1	2	8	6	0.284	2
ZUPCO minibuses (ZM)	1/3	1/2	1	7	5	0.192	3
Non-ZUPCO omnibuses (NZO)	1/9	1/8	1/7	1	1/3	0.032	4
Pirating small cars (PSC)	1/7	1/6	1/5	3	1	0.06	5
Consistency Ratio = 0.06 < 10%, Acceptable							

Table 8. Alternatives pairwise comparisons with respect to social distancing.

Social distancing (SD)	ZB	ZO	ZM	NZO	PSC	Priority	Rank
ZUPCO buses (ZB)	1	1/2	1/3	2	4	0.16	3
ZUPCO omnibuses (ZO)	2	1	1/2	3	6	0.27	2
ZUPCO minibuses (ZM)	3	2	1	4	8	0.43	1
Non-ZUPCO omnibuses (NZO)	1/2	1/3	1/4	1	2	0.09	4
Pirating small cars (PSC)	1/4	1/6	1/8	1/2	1	0.05	5
Consistency Ratio = 0.03 < 10%, Acceptable							

the lowest (0.06). ZUPCO minibuses scored the highest weight (0.43) whilst pirating small cars have the least (0.05) in relation to observing social distancing. Highest score in terms of temperature screening was scored by ZUPCO-minibuses (0.45) yet the least was (0.04) scored by both pirating small cars and Non-ZUPCO omnibuses operators. In terms of disinfecting, ZUPCO minibuses have the highest score (0.426) whilst pirating small cars and Non-ZUPCO omnibuses scored (0.032) which is the least.

Overall preferences (Composite weights) in the ranking of transport operators in terms of risk minimization on passenger to passenger transmission of COVID-19 virus

Table 9. Alternatives pairwise comparisons with respect to temperature screening

Temperature screening (TS)	ZB	ZO	ZM	NZO	PSC	Priority	Rank
ZUPCO buses (ZB)	1	1/3	¼	6	6	0.16	3
ZUPCO omnibuses (ZO)	3	1	½	8	8	0.30	2
ZUPCO minibuses (ZM)	4	2	1	9	9	0.45	1
Non-ZUPCO omnibuses (NZO)	1/6	1/8	1/9	1	1	0.04	4
Pirating small cars (PSC)	1/6	1/8	1/9	1	1	0.04	4
Consistency Ratio = 0.06 < 10%, Acceptable							

Table 10. Alternatives pairwise comparisons with respect to disinfecting

Disinfecting (D)	ZB	ZO	ZM	NZO	PSC	Priority	Rank
ZUPCO buses (ZB)	1	½	1/3	9	9	0.21	3
ZUPCO omnibuses (ZO)	2	1	1/2	9	9	0.29	2
ZUPCO minibuses (ZM)	3	2	1	9	9	0.426	1
Non-ZUPCO omnibuses (NZO)	1/9	1/9	1/9	1	1	0.032	4
Pirating small cars (PSC)	1/9	1/9	1/9	1	1	0.032	4
Consistency Ratio = 0.09 < 10%, Acceptable							

after considering all the criterions (COVID-19 transmission regulations) are shown in Table 11. The results showed that pirating small cars have the highest composite weight (31.16%), followed by ZUPCO minibuses (29.15%), ZUPCO omnibuses (17.49%), ZUPCO buses (16.98%) and Non-ZUPCO omnibuses (5.23%). In terms of criterion pairwise comparisons, proper mask wearing has the highest score (0.4908), followed by social distancing (0.2566), sanitizer administration (0.1352), disinfecting (0.0802) and temperature screening (0.037). Since AHP pairwise comparisons are based on the user's preference, the researchers valued Proper mask wearing as the most important COVID-19 transmission prevention measure in relation to other measures. This is because, proper mask wearing restricts the virus within the host and in the event that the virus mange to escape and is now circulating in the air, chances of it infecting other passengers are minimized. Social distancing minimizes transfer of the virus by physical contact, sneezing and coughing, among others. This add value to the importance of Proper mask wearing as it ensures safety where social distancing is violated. Sanitization is important in destroying virus in case one has through hands interfered with an infected surface, items among other objects. However, hand sanitization without observing social distancing and proper mask wearing leaves more room for the virus to spread. Disinfecting has the importance of destroying virus in air and on surfaces an effect which is more similar to that of Sanitization though with lesser impact. Temperature screening is important in alarming the operators of the possibility that the passenger has COVID-19 symptoms

Table 11. Overall alternative composite weights

	PMW	SD	SA	DIS	TS	Composite weight	Rank
Criterion overall weights	**0.4908**	**0.2566**	**0.1352**	**0.0802**	**0.037**		
ZUPCO buses (ZB)	0.1	0.16	0.432	0.16	0.21	0.169(16.98%)	4
ZUPCO omnibuses (ZO)	0.064	0.27	0.284	0.3	0.29	0.174(17.49%)	3
ZUPCO minibuses (ZM)	0.208	0.43	0.192	0.45	0.426	0.290(29.15%)	2
Non-ZUPCO omnibuses (NZO)	0.042	0.09	0.032	0.04	0.032	0.052(5.23%)	5
Pirating small cars (PSC)	0.58	0.05	0.06	0.04	0.032	0.310(31.16%)	1

(a body temperature of above 37 °C). This however has the limitation when it comes to asymptomatic individuals and those suffering from other conditions not specifically COVID-19. Temperature screening also comes late for preventing the spreading of the virus where social distancing and proper mask wearing are violated making it to be of lesser importance. This again points to the importance of social distancing and proper mask wearing in containing the virus.

6 Conclusions

AHP has proved to be a versatile methodology in dealing with multi-criteria decision making problems in the fields of transport and public health. From this study, policy makers and law enforcement agents are hinted grey areas in relation to which transport operator category is lagging behind in terms of COVID-19 prevention measures and their implementation. On the other hand public transport users are sensitized on the need for making informed decision on which operator category has the least risk of them contracting the so deadly and infectious Corona virus as they travel from one place to the other.

Conflicts of Interest. The authors have no conflict of interest to declare that are relevant to the content of this manuscript.

References

1. Mohammed, H.J., Daham, H.A.: Analytic hierarchy process for evaluating flipped classroom learning. Comp. Mat. Contin. **66**, 2229–2239 (2021)

2. Briceño-León, C.X., Sanchez-Ferrer, D.S., Iglesias-Rey, P.L., Martinez-Solano, F.J., Mora-Melia, D.: Methodology for pumping station design based on analytic hierarchy process (AHP). Water **13**, 2886 (2021)
3. Zekhnini, K., Cherrafi, A., Bouhaddou, I., Benghabrit, Y.: Analytic hierarchy process (AHP) for supply chain 4.0 risks management. In: Artificial Intelligence and Industrial Applications: Smart Operation Management, pp. 89–102 (2021)
4. Dar, T., Rai, N., Bhat, A.: Delineation of potential groundwater recharge zones using analytical hierarchy process (AHP). Geol., Ecol., Lands. **5**, 292–307(2021)
5. Hussain, S., Xuetong, W., Hussain, T., Khoja, A.H., Zia, M.Z.: Assessing the impact of COVID-19 and safety parameters on energy project performance with an analytical hierarchy process. Utilit. Pol. **70**, 101210 (2021)
6. Saha, J., Haldar, S., Bhattacharya, S., Paul, S.: Tourism in retrospect of COVID-19 on global perspective using analytical hierarchy process. Spat. Inf. Res. **29**(6), 981–995 (2021). https://doi.org/10.1007/s41324-021-00407-4
7. Dewi, N.K., Putra, A.S.: Decision support system for head of warehouse selection recommendation using analytic hierarchy process (AHP) method. In: International Conference Universitas Pekalongan, vol. 1, pp. 43–50 (2021)
8. Hermawan, R., Habibie, M.T., Sutrisno, D., Putra, A.S., Aisyah, N.: Decision support system for the best employee selection recommendation using Ahp (Analytic Hierarchy Process) method. Int. J. Educ. Res. Soc. Sci. **2**, 218–1226 (2021)
9. Valentino, V.H., Setiawan, H.S., Saputra, A., Haryanto, Y., Putra, A.S.: Decision support system for thesis session pass recommendation using AHP (analytic hierarchy process) method. Int. J. Educ. Res. Soc. Sci. **2**, 215–221 (2021)
10. Chaudhary, S., Kumar, A., Pramanik, M., Negi, M.S.: Land evaluation and sustainable development of ecotourism in the Garhwal Himalayan region using geospatial technology and analytical hierarchy process. Environ. Dev. Sustain. **24**(2), 2225–2266 (2021). https://doi.org/10.1007/s10668-021-01528-4
11. Naseer, S., Song, H., Aslam, M.S., Abdul, D., Tanveer, A.: Assessment of green economic efficiency in China using analytical hierarchical process (AHP). Soft. Comput. **1**, 11 (2021). https://doi.org/10.1007/s00500-021-06507-5
12. Naveed, Q.N., Qahmash, A.I., Al-Razgan, M., Qureshi, K.M., Qureshi, M.R.N.M., Alwan, A.A.: Evaluating and prioritizing barriers for sustainable e-learning using analytic hierarchy process-group decision making. Sustainability **14**, 8973 (2022)
13. Fam, S.F., Chuan, Z.L., Wahjono, S.I., Zaini, N.D.: Fuzzy analytical hierarchy process (F-AHP) method in evaluating e-wallet payment system in Malaysia. Math. Statist. Eng. Appl., 742–749 (2022)
14. Liu, Z., Ma, R., Wang, H.: Assessing urban resilience to public health disaster using the rough analytic hierarchy process method: a regional study in China. J. Safety Sci. Resilience **3**, 93–104 (2022)
15. Klumpp, M., Monfared, I.G., Vollmer, S.: Public opinion on global distribution of COVID-19 vaccines: evidence from two nationally representative surveys in Germany and the United States. Vaccine **40**, 2457–2461 (2022)
16. Kumar, V., Nguyen, H.T.T., Mittal, A., Lai, K.K.: The production and distribution of face masks to other countries: a strategic approach of Taiwan during COVID-19 outbreak. J. Global Operat. Strategic Sourcing (2022)
17. Gupta, R., Rathore, B., Srivastava, A., Biswas, B.: Decision-making framework for identifying regions vulnerable to transmission of COVID-19 pandemic. Comput. Ind. Eng. **169**, 108207 (2022)

Green Task Scheduling Algorithm in Green-Cloud

Mysha Maliha Priyanka[1], K. M. Safin Kamal[1], Md. Hasibur Rahman[1], Ayesha Siddiqa Yasmin[1], Ahmed Wasif Reza[1](✉), and Mohammad Shamsul Arefin[2](✉)

[1] Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh
wasif@ewubd.edu

[2] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattoram 4349, Bangladesh
sarefin@cuet.ac.bd

Abstract. Cloud-dedicated servers can better meet green computing standards by being ecologically friendly. "Green cloud computing" refers to utilizing information technology and other technological achievements to help the environment. Task scheduling is one of the biggest challenges in cloud-based systems that must be addressed to improve system efficacy and user experience. The primary goal of this study is to develop an algorithm that focuses on minimizing green cloud computing execution times while remaining environmentally friendly. We compared task scheduling algorithms based on execution time, such as FCFS, SJF, and Round Robin, to the approach we recommended, the generalized priority (GP) algorithm. We experimented with evaluating our technique using the CloudSim 3.0.3 simulator. The algorithm we suggested has the shortest runtime out of all the algorithms which is 97.91.

Keywords: Cloud computing · Virtual machine (VM) · Task scheduling algorithm · CloudSim 3.0.3 simulator

1 Introduction

"Cloud Computing" refers explicitly to everything that involves hosting services online. There are three types of cloud services, such as IaaS, PaaS, and SaaS [1]. For clients, the services are pay-as-you-go or available as needed. These services, primarily global and diversified, can be deployed and released quickly with little management work or managed service involvement. A cloud-based data center with additional servers is typically deployed. Regarding overall expense and usage, cloud computing has several benefits. The cloud service provider manages values provided by data users [2]. Using technology innovations like computing and other IT services in a sustainable way for significant environmental advantages is known as "Green Cloud Computing."

A computer program's ability to do work in a particular order is controlled by a set of standards known as scheduling [3]. Among all the algorithms for scheduling, there

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 75–87, 2024.
https://doi.org/10.1007/978-3-031-50151-7_8

is a task scheduling algorithm. The primary benefit of the task scheduling technique is the optimum system efficiency and high computing performance. We can employ task scheduling methods to cut costs and save time. In this paper, the essential facts are,

- We have proposed an algorithm called the Generalized Priority (GP) Algorithm.
- We have compared some algorithms like First Come First Serve (FCFS), Round Robin (RR), Shortest Job First (SJF) algorithm, and GP algorithm to find the most efficient algorithm which minimizes the time.

The content of this paper is as described in the following: in Sect. 1, the basic introduction of the cloud computing and task scheduling algorithm is given. Then, Sect. 2 is about the related work on the same topic. After that, in Sect. 3, the research methodology is discussed. The results and discussion are in the 4th section. Finally, in Sect. 5, the paper concludes.

2 Related Work

Ismail and Materwala in 2018 [1] proposed a task scheduling algorithm which is called "EATSVM (Energy-aware task scheduling on Cloud Virtual Machines)." The paper formed a cloud-based computing system with five different types of virtual machines which optimizes the energy in a heterogeneous cloud method. The research claimed that the "EATSVM algorithm" conserves more energy than their compared "ECTC algorithm."

In [2], Kak et al. summarized the relevant work on the methods of the energy-efficient task scheduling methods that are now in use. In order to attain the goal of decreased energy usage and CO_2 emission in a cloud environment, they have presented a model. Machine learning technologies have been used in this paper. The objective of this paper was to reduce makespan and energy consumption, maximize resource utilization and minimize the execution time. The model will eventually be implemented within MATLAB and evaluated on several factors, including makespan, execution time, resource use, QoS, and energy usage.

In 2022, Mansouri, Najme, and Ghafari, R. proposed a Cost-Efficient Task Scheduling Algorithm (CETSA) [3]. The primary goal was to reduce makespan, energy expenditure, and cost in cloud computing. The MSDE, CPSO, CJS, and FUGE algorithms were compared with the CETSA algorithm using the CloudSim simulator.

In [4], the algorithms used are "Green cloud scheduling algorithm" and compared with other algorithms for task schedule such as HEROS scheduler, Random scheduler, and round robin algorithm. Rao, G. J. and Babu, G. S. used the Green cloud simulator to compare the algorithms. The main intention of the paper was to reduce carbon emissions and energy efficiency.

In 2016, Maharana, D., Sahoo, D. B. and Sethi, D. S. suggested EDVS/dynamic energy-efficient real-time tasks scheduling, assessed using the CloudSim toolkit [5]. The primary goals of the paper are to shrink energy usage in cloud-based data centers and enhance system performance. The work in [17–26] shows different guidelines forgreen ICT.

3 Research Methodology

Networking, parallel, distributed, and cloud computing performance has all been significantly impacted by resource allocation and scheduling. Several researchers have proposed various algorithms for effective scheduling, assignment, and scaling of all the given resources in the cloud. However, here, we are discussing four scheduling algorithms.

1. First come first serve
2. Round Robin
3. Shortest job first scheduling and
4. Generalized Priority Algorithm (GPA) is our proposed algorithm.

3.1 Description of the Algorithms

Here we are showing the description of algorithms and their flowchart and procedure.

- **FCFS (First Come First Serve)**

FCFS is a well-known task-scheduling algorithm that chooses the resources for incoming requests for parallel processing. The algorithm, FCFS's complete form is First Come First Serve, and it works similarly. The process which comes first, the algorithm schedules it first. It is the one with the shortest wait time [6]. The cloud sim provides FCFS scheduling techniques and internal task scheduling. A cloud-based data center's virtual machine (VM) components are responsible for allocating app-specific VMs to the hosts. However, FCFS has a very slow turnaround and response time [7]. However, as the algorithm schedules like the non-preemptive method, the next task will start after finishing a complete task. As a result, FCFS may have gotten into a starvation problem. The shortest jobs at the back of the queue must wait for an extended period because they must wait until the front of the queue is completed [8]. In Fig. 1, we show the flowchart of FCFS. Here we can see how the algorithm works.

- **Round Robin (RR) Algorithm**

The Round Robin algorithm is one of the most popular task-scheduling algorithms. It gives a starvation-free task schedule and also has a straightforward and simple implementation. The Round Robin algorithm is named after a well-known principle called the 'Round Robin Principle.' Which stands for, turn by turn, each person will get an equal-weighted share of something. The cloud Sim provides RR scheduling techniques and internal task scheduling. A cloud-based data center's VM components allocate app-specific VMs to the hosts. It performs the time allocation without assigning priority rather than handles each task in an identical proportion. This algorithm is better than FCFS [9]. Figure 2 shows the flowchart of RR. Here we can see how the algorithm works.

- **SJF (Shortest Job First) Algorithm**

Shortest Job First, or SJF, is another well-known algorithm. SJF term stands for short job. It works like its title: selecting the waiting request or process with a short execution time means the waiting process with the shortest execution time will be scheduled and

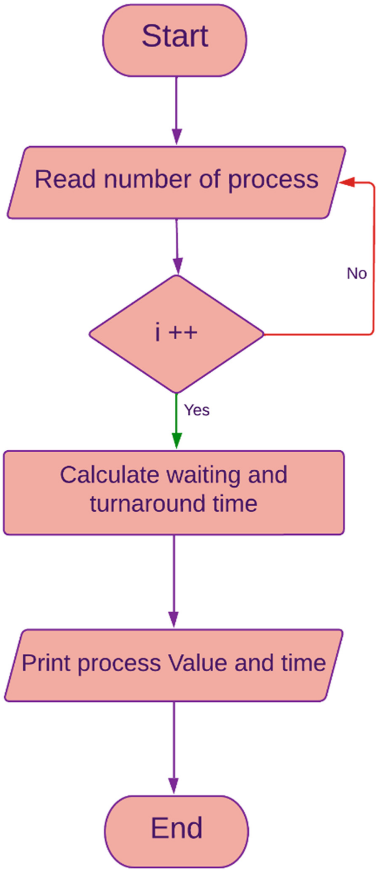


Fig. 1 Flowchart of FCFS

executed first. The cloud Sim provides SJF scheduling techniques and internal task scheduling. We usually use this algorithm for long schedules. SJF algorithm is of two types [8, 10]. It can be worked preemptively and non-preemptively. However, SJF can cause long-term starvation problems as it has a long turn-around (TA) time. Also, it must maintain the average turn-around time [11]. In Fig. 3, we are showing the flowchart. Here we can see how the algorithm works.

- **Generalized Priority (GP) Algorithm**

The customer establishes the priority following user demand, and one must provide the cloudlet's parameters, such as bandwidth size of the cloudlets, memory, and scheduling rules. The given tasks are initially ranked in the suggested method according to their size, with the work with the largest size receiving the top position. The MIPS value of each Virtual Machine is also used to rank (prioritize) them, with the highest MIPS having the highest rank. As a result, the work size and the VM's MIPS are important factors in assigning tasks a priority. In comparison to FCFS and SJF, this approach performs

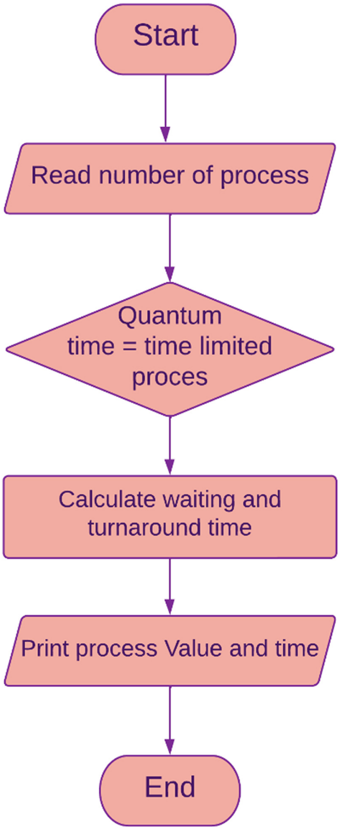


Fig. 2 Flowchart of RR

better. This is our proposed algorithm. In Fig. 4, the algorithm procedure is given. Here we can see how the algorithm works.

Algorithm Procedure:

```
Begin
Push first request or node
While Stack is not Empty do
    Obtain the unvisited request adjacent next to the Stack top
    if no adjacent no then
        If previous = Top on the stack
        Insert all contents of stack to VM list
        End if
    End if
    Pop
    if Stack = Empty then
        Previous = Top on the stack
    else
        The node or request is visited
        Push adjacent requests
    End if
End
```

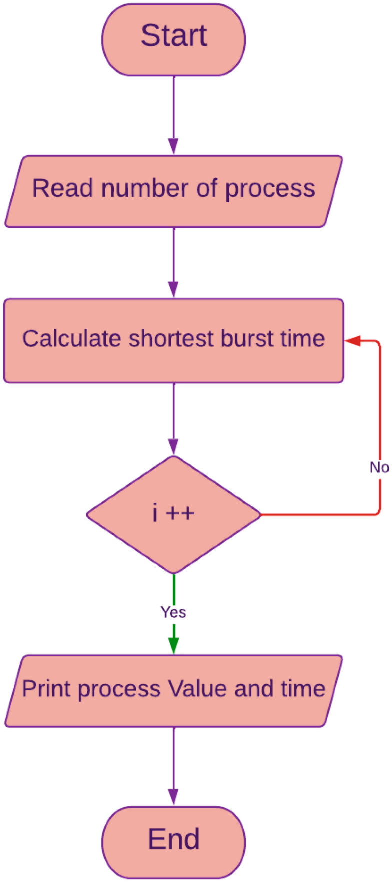


Fig. 3 Flowchart of SJF

3.2 Simulation Environment Setup

For the testing, we have to create an environment for simulation. Here in Figs. 5 and 6, we are showing a framework and how task scheduling works.

Users send their requests to the data center where many data servers exist. When there are many requests, the task scheduler sends them according to its algorithm [12].

Based on the outcomes of the simulation, the efficiency of the task scheduling algorithms is evaluated. The CloudSim 3.0.3 simulator was used for the experiment. One of the most well-liked simulators for the cloud data center is CloudSim. The performance of different task-scheduling algorithms can be compared using this simulator [13]. Figure 7 shows the structure of the CloudSim. The following instructions are for operating it:

- Launch a JAVA IDE and initialize the CloudSim package.
- To organize and distribute resources, establish a data center and broker.

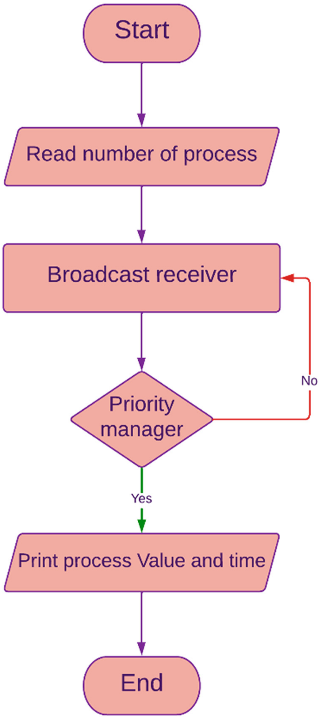


Fig. 4 Flowchart of GPA

- Create virtual machines (VMs), add them to a list, and then transmit the list to the data center broker.
- Create a cloud task set. After creating a cloud task set and including the cloud task cloudlet in it, submit the cloud task set to the broker.
- Start the simulation, and once it is finished, print the results.

In Fig. 7, there are a total of three layers in the CloudSim. In the first layer, users insert their codes. The second layer, like VM service, Cloud service, cloud resource, and network, controls critical element execution. Lastly, the third layer is the core simulation engine of CloudSim [14, 15].

There are other parameters in cloudsim, like the number of VMs, cloudlets, MIPS, and others, that we had to initiate [16]. Table 1 shows the parameters and values we have used for our research in the simulation environment for the experiments:

We have set up the simulation environment on our Intel i5 7th Gen 4Core computer with 16GB RAM and an NVIDIA GTX 1060 GPU.

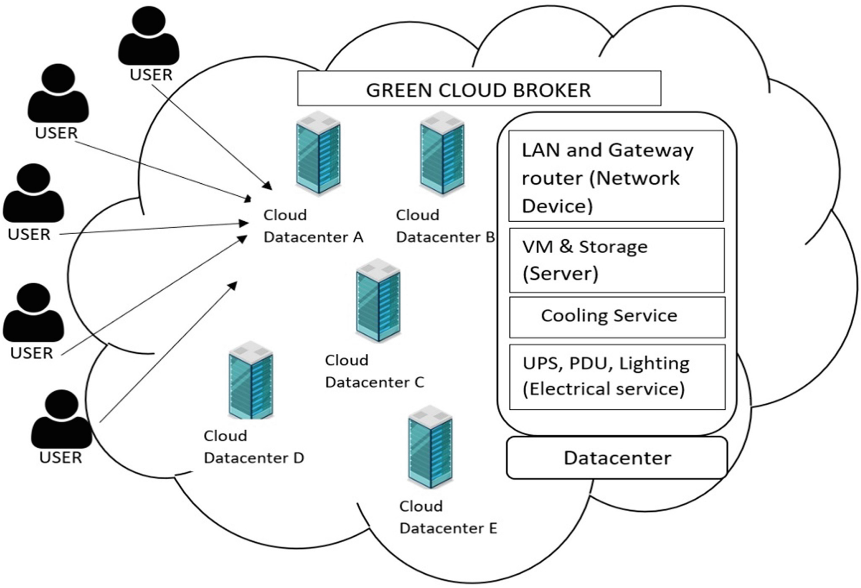


Fig. 5 Cloud framework

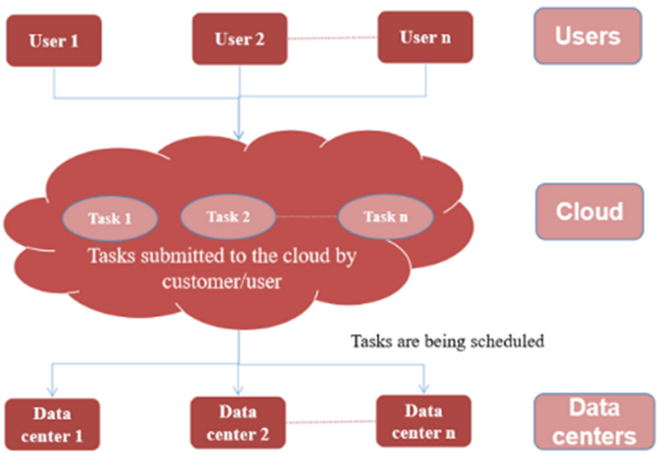


Fig. 6 Task scheduling in cloud

4 Results and Discussion

After setting up the CloudSim environment, we simulated the task scheduling algorithm. In the simulation, we have got execution times. In Table 2, we are trying to compare four algorithms: FCFS, SJF, Round Robin, and the Generalized priority algorithm. Here, we used 13 cloudlets and 100 VMs randomly.

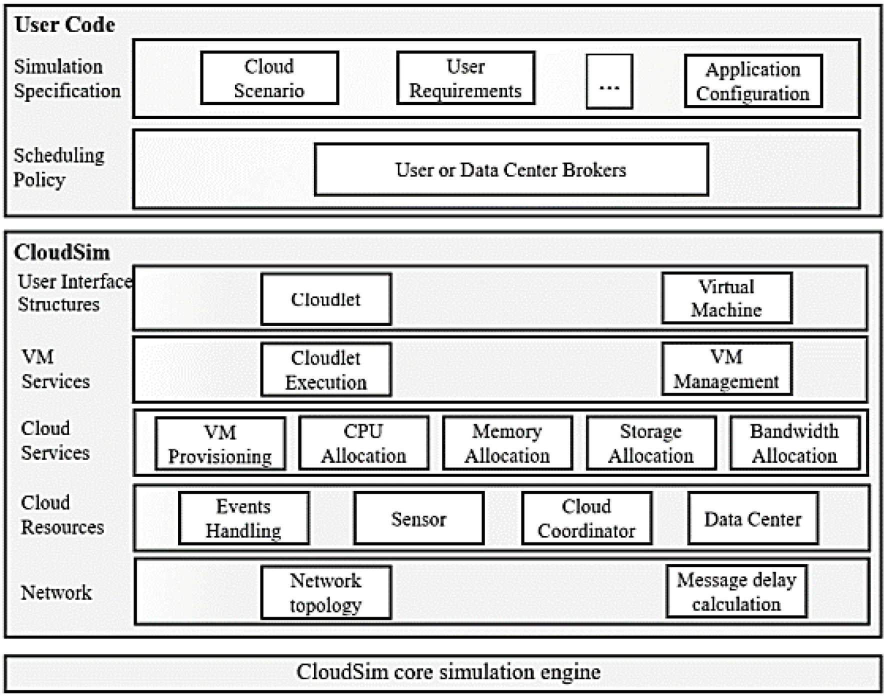


Fig. 7 CloudSim structure

Average Time. In Table 3, we compared the average time of four algorithms which is FCFS, SJF, Round Robin, and Generalized Priority Algorithm.

Figure 8 shows the task scheduling algorithm's performance according to time.

FCFS algorithm gives different values for cloudlet and VM IDs. We took 13 cloudlets and 100 VMS to measure the performance. Time started from 41.83 s for the 1st data, and for the 13th data, it is 150.63 s. In the case of SJF, Time started from 41.83 s for the 1st data, and for the 13th data, it is 152.91 s. For Round Robin, Time started from 41.80 s for the 1st data, and for the 13th data, it is 145.63 s. Lastly, our proposed one GP algorithm Time started from 32.71 s for the 1st data, and for the 13th data, it is 105.94 s. Our generalized priority algorithm works better if we see the graphs (Figs. 8 and 9) and analyze them. It takes less time, so its energy consumption is lower than others. Considering the average time of these four algorithms, we get better results for the Generalized Priority algorithm.

Table 1 Parameter and values for CloudSim environment

Parameters	Value
Number of data centers	1
Operating system	Linux
Total number of VMs	100
MIPS of processing element	250
VM memory (RAM)	512MB
Number of CPUs of VM	1
Tasks size of VM	1000MB
Number of cloudlets	12–100
Length of cloudlets	4000
File size	300
Host memory (MB)	102400 MB
Host storage	1TB
Time zone	10.0
Processing cost	0.3$/Core (hourly)
Memory cost	0.05$/GB (hourly)
Storage cost	0.001$/GB (monthly)

Table 2 Output for all algorithm in data center ID 2

FCFS			SJF			Round Robin algorithm			Generalized priority algorithm		
Cloudlet ID	VM	Time	Cloudlet ID	VM	Time	Cloudlet ID	VM	Time	Cloudlet ID	VM	Time
0	0	41.83	0	0	41.83	0	0	41.80	0	0	32.71
1	1	68.55	1	1	68.55	1	1	67.55	1	1	52.63
2	2	82.2	2	2	79.93	2	2	82.1	2	2	62.22
3	4	92.88	3	4	92.15	3	4	90.88	3	4	71.43
5	3	96.09	5	3	95.31	5	3	91.09	5	3	71.7
4	6	104.8	4	6	103.9	4	6	103.8	7	5	80.25
7	5	110.2	7	5	109.3	7	5	108.2	4	6	80.57
6	8	116.5	6	8	115.6	6	8	111.5	6	8	88.86
9	7	124.6	9	7	123.5	9	7	120.5	9	7	89.23
8	10	128.1	8	11	127.1	8	12	125.1	8	11	97.27
11	9	139.1	11	9	139.1	11	9	138.2	11	9	97.68
10	11	139.4	10	10	137.98	10	10	137.43	10	10	105.48
13	13	150.63	13	12	152.9	13	11	145.63	13	13	105.94

Table 3 Average time of all algorithms

FCFS	SJF	Round Robin (RR)	Generalized priority (GP)
131.53	131.98	104.90	97.91

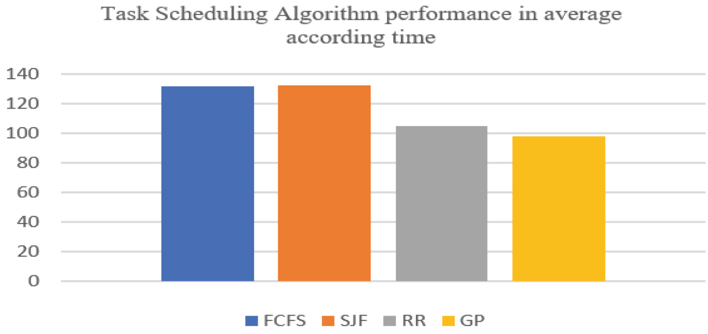


Fig. 8 Bar chart of the average times

5 Conclusion

One of the most crucial duties in a green cloud computing environment is scheduling. When a task takes less time to execute, it will help to carbon emission less, which will affect our environment by making it green. Various scheduling algorithms that effectively plan out computational work in a green-cloud setting have been examined in this research. We tested the FCFS, SJF, and round-robin scheduling algorithms and suggested the GPA (Generalized Priority Algorithm) as a novel scheduling technique. In cloud systems, priority is a key factor in job scheduling. The experiment is run with various workload traces and virtual machine counts. Round Robin, SJF, and FCFS are contrasted with the experiment undertaken. The outcome demonstrates that the suggested algorithm is much more effective than FCFS, SJF, and Round Robin. In this paper, we primarily analyze four algorithms. We proposed an algorithm that is a generalized priority task scheduling algorithm using a small number of tasks. We have a target in the future that we will add some more tasks and attempts like ML, AI, and deep learning to shorten the time needed for execution. We also propose this algorithm for a big grid ecosystem and will compare the times that need for execution between the grid and the green cloud.

The need for more effective techniques to plan cloud computing jobs that reduce energy usage and carbon dioxide emission is the research gap in green task scheduling algorithms in green clouds. Traditional task scheduling algorithms waste a lot of energy and cause environmental degradation, according to numerous studies. Green task scheduling algorithms work to address these problems by employing creative methods that consume less energy and leave less of a carbon impact. As the world's reliance on cloud computing grows, it is essential to create more environmentally friendly and sustainable procedures to make sure that the industry's expansion does not harm the environment. It can be used at the industrial level.

References

1. Ismail, L., Materwala, H.: EATSVM: Energy-aware task scheduling on cloud virtual machines. Proc. Comput. Sci. **135**, 248–258 (2018). Available at: https://doi.org/10.1016/j.procs.2018.08.172
2. Kak, S.M., Agarwal, P., Alam, M.A.: Task scheduling techniques for energy efficiency in the cloud. EAI Endorsed Transac Energy Web **9**(39) (2022). https://doi.org/10.4108/ew.v9i39.1509
3. Mansouri, N., Ghafari, R.: Cost-efficient task scheduling algorithm for reducing energy consumption and makespan of cloud computing, 5, pp. 1–10 (2022). Available at: https://doi.org/10.22067/CKE.2022.70000.1008
4. Rao, G.J., Babu, G.S.: Energy analysis of task scheduling algorithms in Green Cloud. In: 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) [Preprint] (2017). Available at: https://doi.org/10.1109/icimia.2017.7975624
5. Maharana, D., Sahoo, D.B., Sethi, D.S.: Energy-Efficient Real-Time Tasks Scheduling in Cloud Data Centers, pp. 768–773. https://core.ac.uk/download/pdf/235196825.pdf (no date)
6. Bhat, R.V. et al.: Survey on Energy and Cost Optimization Techniques using Green Cloud computing **XVI**(V), pp. 337–344 (2020)
7. Siahaan, A.P.: Comparison Analysis of CPU Scheduling FCFS, SJF and Round Robin (2017). Available at: https://doi.org/10.31227/osf.io/6dq3p
8. Mhedheb, Y., Streit, A.: Energy-efficient task scheduling in data centers. In: Proceedings of the 6th International Conference on Cloud Computing and Services Science [Preprint] (2016). Available at: https://doi.org/10.5220/0005880802730282
9. Ismail, L., Fardoun, A.A.: Towards energy-aware task scheduling (EATS) framework for divisible-load applications in cloud computing infrastructure. In: 2017 Annual IEEE International Systems Conference (SysCon) (2017) https://doi.org/10.1109/syscon.2017.7934791
10. Greenberg, S., Mills, E., Tschudi, B.: Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers. SUMMER, 76–87 (no date)
11. Maharana, D., Sahoo, D.B., Sethi, D.S.: Energy-Efficient Real-Time Tasks Scheduling in Cloud Data Centers, pp. 768–773 (2016)
12. Gade, A.: Survey on energy efficient cloud: A novel approach towards green computing. HELIX **8**(5), 3976–3979 (2018). Available at: https://doi.org/10.29042/2018-3976-3979
13. Yao, K.R.A., Francois, M. et al.: Energy Efficiency and Scheduling Techniques Using Cloud Computing Methods Within Data Centers, 97 (2019)
14. Jena, S.R., Shanmugam, R., Saini, K., Kumar, S.: Cloud computing tools: inside views and analysis. Proc. Comput. Sci. **173**, 382–391 (2020) (Between cloud)
15. Hamann, H.F.: Optimization Algorithms for Energy-Efficient Data Centers, Volume 2: Thermal Management; Data Centers and Energy Efficient Electronic Systems [Preprint] (2013). Available at: https://doi.org/10.1115/ipack2013-73066
16. Atiewi, S., Yussof, S., Ezanee, M.: A comparative analysis of task scheduling algorithms of virtual machines in cloud environment. J. Comput. Sci. **11**(6), 804–812 (2015). Available at: https://doi.org/10.3844/jcssp.2015.804.812
17. Yeasmin, S., Afrin, N., Saif, K., Reza, A.W., Arefin, M.S.: Towards building a sustainable system of data center cooling and power management utilizing renewable energy. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_67
18. Liza, M.A., Suny, A., Shahjahan, R.M.B., Reza, A.W., Arefin, M.S.: Minimizing e-waste through improved virtualization. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A.,

Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_97
19. Das, K., Saha, S., Chowdhury, S., Reza, A.W., Paul, S., Arefin, M.S.: A sustainable e-waste management system and recycling trade for Bangladesh in Green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_33
20. Rahman, M.A., Asif, S., Hossain, M.S., Alam, T., Reza, A.W., Arefin, M.S.: A sustainable approach to reduce power consumption and harmful effects of cellular base stations. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_66
21. Ahsan, M., Yousuf, M., Rahman, M., Proma, F.I., Reza, A.W., Arefin, M.S.: Designing a sustainable e-waste management framework for Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_104
22. Mukto, M.M., Al Mahmud, M.M., Ahmed, M.A., Haque, I., Reza, A.W., Arefin, M.S.: A sustainable approach between satellite and traditional broadband transmission technologies based on Green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_26
23. Meharaj-Ul-Mahmmud, Laskar, M.S., Arafin, M., Molla, M.S., Reza, A.W., Arefin, M.S.: Improved virtualization to reduce e-waste in green computing. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_35
24. Banik, P., Rahat, M.S.A., Rafe, M.A.H., Reza, A.W., Arefin, M.S.: Developing an energy cost calculator for solar. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_75
25. Ahmed, F., Basak, B., Chakraborty, S., Karmokar, T., Reza, A.W., Arefin, M.S.: Sustainable and profitable IT infrastructure of Bangladesh using Green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_18
26. Ananna, S.S., Supty, N.S., Shorna, I.J., Reza, A.W., Arefin, M.S.: A policy framework for improving e-waste management in Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_95

Estimation of Optimum Design of a 3-Bar Truss System with Decision Tree Algorithm

Ayla Ocak[1], Gebrail Bekdaş[1], Ümit Işıkdağ[2], and Sinan Melih Nigdeli[1](✉)

[1] Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcılar/Istanbul, Turkey
melihnig@iuc.edu.tr
[2] Department of Informatics, Mimar Sinan Fine Arts University, 34427 Şişli/Istanbul, Turkey

Abstract. Truss systems are structures that require care and cost due to the materials and workmanship used. In the design of such systems, the process of optimizing the material volume is necessary for optimum cost. In terms of convenience in application, the production of materials in standard sections also reduces labor costs. In this study, the cross-sectional areas of the bars were optimized to minimize the volume for the design of a 3-bar truss system. Harmony Search Algorithm (HSA), a metaheuristic algorithm inspired by nature, was used in the optimization. A data set was prepared by determining the optimum cross-sectional areas for certain load and stress ranges, and a machine learning prediction model based on the load and stress information with the decision tree classification algorithm was produced. For this purpose, the bar cross-section areas in the data were converted to standard cross-sections and divided into classes. With the produced model, under the desired load and stress values, the bar cross-sectional areas of the system were estimated on a class basis. When the results were examined, it was determined that the prediction model produced with the optimum data was successful at a level of approximately 95% in estimating the bar sections.

Keywords: Truss system · Optimization · Metaheuristic algorithm · Harmony search algorithm · Machine learning · Decision tree classification

1 Introduction

Truss systems are civil engineering structure elements that require good material workmanship. In the design of these systems, the main parameters considered are stress and external loads on the system. The overall durability of the truss system against external loads is provided by the cross-section dimensions. The production, transportation, and placement of the materials used in the truss systems constitute a large part of the construction cost. Nowadays, the optimization process is frequently used to reduce the construction costs of truss-type structures. In this optimization process, minimizing the material volume constitutes the objective function in obtaining the optimum cost. In the optimization of truss systems, metaheuristic algorithms inspired by nature are preferred due to their simple and understandable structure and easy applicability to problems.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 88–97, 2024.
https://doi.org/10.1007/978-3-031-50151-7_9

They contain design factors specific to the natural event or living behavior that they are inspired by. Metaheuristic algorithms are inspired by events such as the pollination process of flowers, bees, and ants following the road for food, bats locating by echolocation, and the teaching and learning process of students and use factors specific to these events [1–5]. In studies with metaheuristic algorithms, metaheuristic algorithms such as the Jaya algorithm, flower pollination algorithm, bat algorithm, gray wolf optimization, artificial bee colony algorithm, firefly algorithm, and genetic algorithm have been used [6–9]. The harmony search algorithm developed by Geem et al. is another metaheuristic algorithm frequently preferred in the optimization of lattice systems [10]. Studies have supported the success of the harmony search algorithm in civil engineering problems [11–16]. The use of this algorithm and its modifications in lattice system studies has become quite common [17–22].

Material dimensions come first among the things that should be considered in the design of truss systems. The production of materials in standard sections both facilitates production and ensures efficient use of the workforce. In today's technology, artificial intelligence technologies have started to be used for optimum production efficiency. In the application of artificial intelligence to such problems, production forecasting models suitable for the purpose can be easily obtained by using the optimization process in data production. Machine learning, a sub-branch of artificial intelligence, is an artificial intelligence method that enables the machine to make inferences from data by training. There are applications of machine learning in lattice system designs. Machine learning methods have been successfully used in various civil engineering truss system problems such as the design of plane trusses, structural damage detection of truss bridges, nut-bolt loss estimation of truss bridges, safety assessment of trusses, and strength estimation of trusses [23–28].

In this study, a 3-bar truss system example is optimized on the Matlab program for various load and stress combinations so that the material volume is minimized [29]. In the optimization, the harmony search algorithm was used, and the bar cross-section areas were optimized. The optimum results obtained were turned into a data set and the classification process was applied with machine learning. In the data set, standard production for bar sections was aimed, and bar areas were converted into classes at certain intervals. With the decision tree algorithm, which is a machine learning algorithm, a model was created in the Python program that predicts the bar sections used in the system [30]. With the produced model, it was provided to estimate the optimum bar sections for various load and stress combinations.

2 Methodology

2.1 Optimal Design of Truss Systems and Prediction with Artificial Intelligence

In the optimization of truss systems, the main purpose is to reduce the material volume. The shortest way to do this is to set the material sections. In the optimization process, optimizing the bar cross-sectional areas makes the truss system designs more economical. Constraints for design are also constraints to limit volume. Optimization of truss systems is based on external loads and stress conditions.

The harmony search algorithm is an algorithm inspired by the process of finding the best harmony. It includes design factors such as harmony memory consideration rate (HMCR) and fret width (FW). These factors take part in deciding the equations to be used in the production of the optimum solution. In optimization with the harmony search algorithm, design constants are introduced to the system like other heuristic algorithms, and lower and upper limits of parameters and algorithm-specific factors are entered as inputs. Then, a random solution vector is generated for the optimized variable within the limits given by the objective function. Equation 1 shows the equation for creating a harmony vector.

$$X_{new} = X_{\min} + rand(1)(X_{\max} - X_{\min}) \tag{1}$$

X_{new} in Eq. 1 represents each new value in the harmony vector, and $X_{\max}$ and $X_{\min}$ represent the upper and lower limit values of the X parameter, respectively.

Another way to create a harmony vector is to use the fret width parameter. The memory consideration ratio (HMCR) decides which equation to use as the harmony vector generation equation. Accordingly, a random value between 0 and 1 is assigned and the solution equation is decided according to whether this value is greater or less than the HMCR value defined at the beginning. The solution equation produced with the fret width parameter is given in Eq. 2.

$$X_{new} = X_n + rand(1)FW(X_{\max} - X_{\min}) \tag{2}$$

The FW given in Eq. 2 is an algorithm-specific constant. X_n value denotes the nth solution of the harmony vector.

Each solution is produced with these two equations in the amount of iteration, and old solutions and new solutions are compared. By updating the better solutions, the harmony vector takes its final form, and the optimum solution is reached.

In the design of truss systems, artificial intelligence methods have opened a suitable field of application. Creating models by performing classification and regression analyses with various artificial intelligence methods contributes to the optimum design of truss systems. One of these methods is machine learning techniques. Machine learning methods, a sub-branch of artificial intelligence, can be applied to truss system problems. In this direction, the basic requirement is to obtain data to be used in machine training. Data sets for the design prediction model to be produced with machine learning can be easily produced thanks to optimization. Classification or regression algorithms are used according to the type of problem in generating a model from the data. The decision tree algorithm is a machine learning algorithm used in classification. It gets its name from the split view of the data used as input, which resembles the branches of a tree when grouping. Each attribute specifies a node point, and these attributes are divided into groups to form tree branches [31]. These branches are separated by the if/else structure and the classification process is completed.

3 The Numerical Example

In this study, a 3-bar truss system sample was taken and an optimization process that minimized the material volume was applied. In the optimization, the cross-sectional areas for the 3-bar system with 2 different types of cross-sections are optimized with the harmony search algorithm in the Matlab program [29]. An application example is shown in Fig. 1.

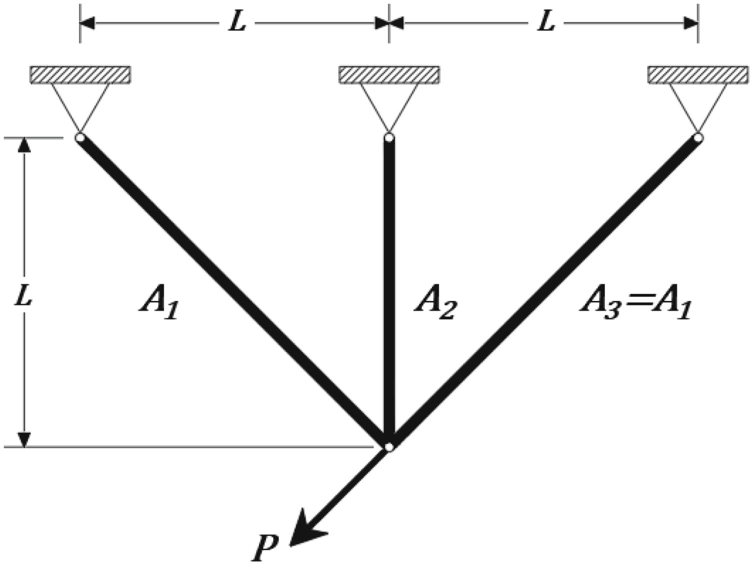


Fig. 1 3-Bar truss system model.

The algorithm parameters used in the optimization process are given in Table 1 where P is the load; sigma is the stress; and A_1 and A_2 are the bar cross-sectional areas. In the design constraints of the 3-bar system, the equations containing the cross-sectional areas related to the load and stress are used. The design constraint equations and the objective function are shown in Table 2.

Table 1 The algorithm parameters

Symbol	Definition	Value
HMCR	Harmony memory considering Rate	0.5
FW	Fret width	0.05
mt	Maximum iteration number	500,000
pn	Population number	15
L	Bar length (cm)	100
P	External load (kN)	1–50
σ	Stress (kN/cm^2)	1–50

Table 2 Design constraint equations and objective functions.

Definition	Equation
Design constraints	$g_1 = \frac{\sqrt{2}A_1 + A_2}{\sqrt{2}A_1^2 + 2A_1A_2}P - \sigma \leq 0$ $g_2 = \frac{A_2}{\sqrt{2}A_1^2 + 2A_1A_2}P - \sigma \leq 0$ $g_3 = \frac{1}{A_1 + \sqrt{2}A_2}P - \sigma \leq 0$
Objective function	Minimize $f(A_1, A_2) = \left(2\sqrt{2}A_1 + A_2\right)L \quad 0 \leq A_1 \leq 1$ and $0 \leq A_2 \leq 1$

As a result of the optimization, 1599 data were generated from different load and stress combinations and converted into a data set to apply the machine learning method. The cross-sectional areas are divided into 10 classes between 0 and 1 cm^2 in terms of standard material production, with 0.1 cm^2 intervals. The first 10 rows of data from the optimum results before classification are shown in Table 3.

Table 3 A part of the dataset generated from optimum results.

Strain (kN/cm^2)	Load (kN)	Volume (cm^3)	A2 (cm^2)	A1 (cm^2)
1	1	263.9012	0.413889	0.7867
2	1	131.9523	0.207586	0.393129
2	2	263.9027	0.415696	0.786066
3	1	87.98576	0.142195	0.260803
3	2	175.9339	0.267498	0.527445
3	3	263.9028	0.41052	0.787897
3	4	353.5782	0.708068	0.999748
4	1	65.9787	0.103106	0.196816
4	2	131.9544	0.20344	0.394602
4	3	197.934	0.316	0.58808

In the created data set, each field is separated into classes by lettering. Stress, load, and volume were used as inputs in the data, and cross-sectional areas were targeted as output. Since there is more than one output, the Multi-Output Classifier method has been applied with various classification algorithms in producing a model that predicts the cross-sectional area of two bars at the same time, and it has been the most successful classifier decision tree algorithm. The comparison of multi-output classification estimation success averages made with various algorithms is given in Table 4.

In multi-output classification, classifications were made that estimate cross-sectional areas separately, considering the success of the decision tree classifier. In classification models of two different types of rods, the dataset is divided into 30% for testing and 70% for training, and the model is subjected to machine learning. The resulting confusion matrices are shown in Fig. 2 for A1 and Fig. 3 for A2. The accuracy, precision, recall, and f1 scores of the models are given in Table 5 for A1 and Table 6 for A2.

Table 4 Multi-output classifier algorithm prediction success table.

Algorithm	Accuracy (%)
Decision tree classification	94.81
K-nearest neighbors classification	93.06
Logistic regression	53.91
Linear discriminant analysis	86.43
Naïve Bayes	88.80
Support vector machine	88.56

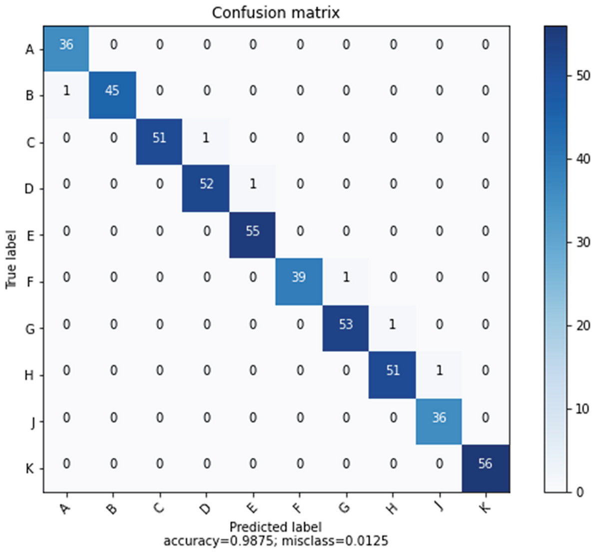


Fig. 2 Confusion matrix for A1 cross-sectional area prediction model.

4 Discussion and Conclusions

In this study, 2 different bar sections used with a harmony search algorithm for cost optimization of a 3-bar system are optimized for different load and stress combinations. The results obtained from the optimization were turned into a data set and a cross-section prediction model based on load and stress was produced by the machine learning

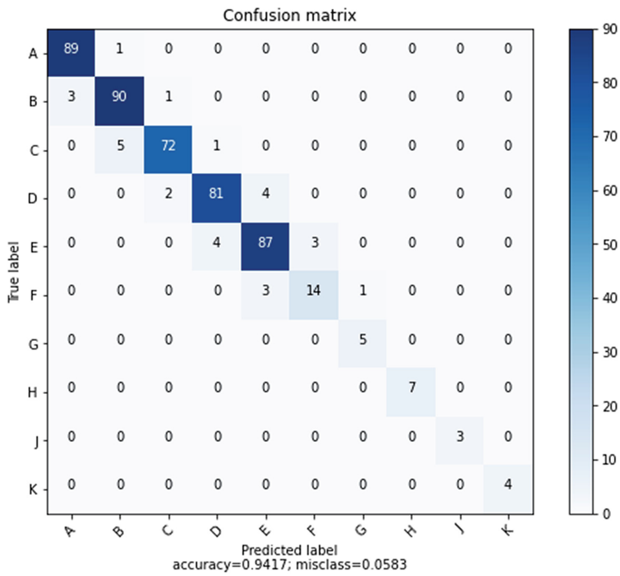


Fig. 3 Confusion matrix for A2 cross-sectional area prediction model.

Table 5 Decision tree classification, accuracy, precision, recall, and f1-score values for A1 cross-section.

Class	Precision	Recall	F1-score
A	0.97	1.00	0.99
B	1.00	0.98	0.99
C	1.00	0.98	0.99
D	0.98	0.98	0.98
E	0.98	1.00	0.99
F	1.00	0.97	0.99
G	0.98	0.98	0.98
H	0.98	0.98	0.98
J	0.97	1.00	0.99
K	1.00	1.00	1.00
Accuracy			0.99

method. In the model that predicts the cross-section of two different bars at the same time, 6 different classification algorithms were tried, and the decision tree classification algorithm provided optimum prediction success. The success rates of the prediction models for the two cross sections produced separately were 94 and 99%. With the model produced for the simultaneous estimation of two cross sections, a prediction success

Table 6 Decision tree classification, accuracy, precision, recall, and f1-score values for A2 cross-section.

Class	Precision	Recall	F1-score
A	0.97	0.99	0.98
B	0.94	0.96	0.95
C	0.96	0.92	0.94
D	0.94	0.93	0.94
E	0.93	0.93	0.93
F	0.82	0.78	0.80
G	0.83	1.00	0.91
H	1.00	1.00	1.00
J	1.00	1.00	1.00
K	1.00	1.00	1.00
Accuracy			0.94

performance of approximately 95% was obtained. In this problem, which is suitable for numerical estimation like the regression, the bar sections are divided into classes and were used in the data for the production of the material in standard molds to reduce the labor costs and production errors, which constitute a significant part of the cost of the truss systems. In the data set containing 10 different classes, a remarkable success level of 95% was found on average. Thus, with the optimization process, while the optimum bar volume is minimized and economic losses are reduced, an optimum design cross-sectional area estimation that minimizes the production errors that may arise in material production with the machine learning method has been made possible with artificial intelligence. In light of all the results, it is understood that it is possible to use optimization in the lattice system design estimation, to increase the reliability and economic performance of the data, and to produce designs that can minimize negative results such as workmanship and production errors by applying artificial intelligence methods.

References

1. Yang, X.S.: Flower pollination algorithm for global optimization, Lecture Notes in Computer Science, In: Durand-Lose, J., Jonoska, N., 27, Springer, London, 7445, 240–249 (2012)
2. Dorigo, M., Maniezzo, V., Colorni, A.: The ant system: an autocatalytic optimizing process. IEEE Trans. Syst. Man Cybernet B **26**, 29–41 (1996)
3. Yang, X.S.: A new metaheuristic bat-inspired algorithm. In: Nature-inspired cooperative strategies for optimization (NICSO 2010), pp. 65–74. Springer, Berlin, Heidelberg (2010)
4. Karaboğa, D.: An idea based on honey bee swarm for numerical optimization, vol. 200, pp. 1–10. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department (2005)

5. Rao, R.V., Savsani, V.J., Vakharia, D.P.: Teaching-Learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput. Aided Des **43**, 303–315 (2011)
6. Bekdaş, G., Yucel, M., Nigdeli, S.M.: Evaluation of metaheuristic-based methods for optimization of truss structures via various algorithms and Lèvy flight modification. Buildings **11**(2), 49 (2021)
7. Miguel, L.F.F., Lopez, R.H., Miguel, L.F.F.: Multimodal size, shape, and topology optimization of truss structures using the firefly algorithm. Adv. Eng. Softw. **56**, 23–37 (2013)
8. Serpik, I.N., Tarasova, N.V.: Optimization of steel trusses with a choice of multi-stage prestressing conditions. Mag. Civil Eng. **96**(5) (2020)
9. Roy, S., Kundu, C.K.: State-of-the-art review on the use of optimization algorithms in steel truss. Int. J. Sci. Technol. Res. **9**(3), 160–165 (2020)
10. Geem, Z.W., Kim, J.H., Loganathan, G.V.: A new heuristic optimization algorithm: harmony search. SIMULATION **76**(2), 60–68 (2001)
11. Siddique, N., Adeli, H.: Applications of harmony search algorithms in engineering. Int. J. Artif. Intell. Tools **24**(06), 1530002 (2015)
12. A novel harmony search-based optimization of reinforced concrete biaxially loaded columns
13. Cakiroglu, C., Islam, K., Bekdaş, G., Kim, S., Geem, Z.W.: Metaheuristic optimization of laminated composite plates with cut-outs. Coatings **11**(10), 1235 (2021)
14. Arama, Z.A., Kayabekir, A.E., Bekdaş, G., Kim, S., Geem, Z.W.: The usage of the harmony search algorithm for the optimal design problem of reinforced concrete retaining walls. Appl. Sci. **11**(3), 1343 (2021)
15. de Almeida, F.S.: Optimization of laminated composite structures using a harmony search algorithm. Compos. Struct. **221**, 110852 (2019)
16. Degertekin, S.O., Minooei, M., Santoro, L., Trentadue, B., Lamberti, L.: Large-scale truss-sizing optimization with enhanced hybrid HS algorithm. Appl. Sci. **11**(7), 3270 (2021)
17. Aydoğdu, A.Y., Artar, M., Ergün, M.: Optimum weight design of steel truss roof systems considering corrosion effect. Structures **49**, 88–105. Elsevier (2023, March)
18. Toklu, Y.C., et al.: Total potential optimization using metaheuristic algorithms for solving nonlinear plane strain systems. Appl. Sci. **11**(7), 3220 (2021)
19. Talatahari, S., Goodarzimehr, V., Taghizadieh, N.: Hybrid teaching-learning-based optimization and harmony search for optimum design of space trusses. J. Optim. Ind. Eng. **13**(1), 177–194 (2020)
20. Cao, H., Qian, X., Zhou, Y.L., Yang, H.: Applicability of subspace harmony search hybrid with improved deb rule in optimizing trusses. J. Comput. Civ. Eng. **32**(4), 04018021 (2018)
21. Artar, M.: A comparative study on optimum design of multi-element truss structures. Steel Compos. Struct. **22**(3), 521–535 (2016)
22. Kaveh, A., Naeimi, M.: Sizing optimization of skeletal structures with a multi-adaptive harmony search algorithm (2015)
23. Zhao, X., Zhang, T., Xiao, W.: An automated design method for plane trusses based on user preference information. Appl. Sci. **13**(3), 1543 (2023)
24. Al-Ghalib, A.A.: Structural damage detection of old ADA steel truss bridge using vibration data. Struct. Control. Health Monit. **29**(11), e3098 (2022)
25. Sun, X., Xin, Y., Wang, Z., Yuan, M., Chen, H.: damage detection of steel truss bridges based on Gaussian Bayesian networks. Buildings **12**(9), 1463 (2022)
26. Li, Z.J., et al.: Detection of nut–bolt loss in steel bridges using deep learning techniques. Sustainability **14**(17), 10837 (2022)
27. Nguyen, T.H., Vu, A.T.: Evaluating structural safety of trusses using machine learning. Frattura ed Integrità Strutturale **15**(58), 308–318 (2021)

28. Truong, V.H., Pham, H.A.: Support vector machine for regression of ultimate strength of trusses: A comparative study. Eng. J. **25**(7), 157–166 (2021)
29. The MathWorks, Matlab R2022b: Natick, MA (2022)
30. Python Software Foundation, 2022, PYTHON 3.11.0, USA
31. Morales, E.F., Escalante, H.J.: A brief introduction to supervised, unsupervised, and reinforcement learning. In: Biosignal Processing and Classification Using Computational Learning and Intelligence, pp. 111–129. Academic Press (2022)

A Soft Computational Technique to Construct a Study Program Recommendation System Based on SDS RIASEC Test

Rolly Intan(✉), Nattayavani Sonia, and Pratiwi Anjasari

Informatics Engineering, Petra Christian University, Surabaya, Indonesia
rintan@petra.ac.id

Abstract. Selection of a program of study based on the talents and interests of students really helps them to be successful in their studies. Here, a psychological test tool is needed to evaluate and correlate student interests and talents with programs of study. Among the psychological test tools, RIASEC is one of the most used psychological test tools to determine suitable careers and programs of studies based on one's interests and talents. This paper discusses and proposes a method using RIASEC dealing with a soft computational technique to build a study program recommendation system for students according to their interests and talents. This recommendation system has been built and implemented at Petra Christian University and received a very positive response from prospective students.

Keywords: Soft computational techniques · Recommendation system · Psychological test · RIASEC

1 Introduction

Prospective students tend to choose a program of study based on external factors, namely by following friends, being determined by their parents, or choosing randomly. Knowing one's own interests and talents (internal factors) will be a better consideration in choosing a program of study. Many students fail to complete their studies because they enter the wrong major that does not match their interests and talents. There are several psychological test tools that can help prospective students choose a major college to match their interests and talents. One of them is RIASEC which stands for *Realistic*, *Investigative*, *Artistic*, *Social*, *Enterprising*, and *Conventional*. RIASEC is a psychological test tool used to determine career and education based on a person's personality. The RIASEC test kit was first introduced by John Holland in the1970s.

In this study, a computer-based system was built to be able to provide recommendations to prospective students in choosing a program of study according to their interests and talents. First, a soft computational technique and method was created and adapted to the concept of the RIASEC Test by making several modifications to accommodate answers from users more smoothly and accurately. The user's answer for each question which is usually answered using two-valued logic (yes or no), is now developed into

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 98–107, 2024.
https://doi.org/10.1007/978-3-031-50151-7_10

7-point Likert scale (*strongly disagree*, *disagree*, *somewhat disagree*, *either agree or disagree*, *somewhat agree*, *agree*, and *strongly agree*). The final goal is to get recommendation results based on the three highest scores on the RIASEC test kit. The three highest scores will produce the appropriate program of studies based on the pattern code that comes out. After getting the three highest scores on the test results, the system will then recommend some program of studies in order based on the highest score in their patterns.

The outline of this paper is as follows. Section 2 will discuss the RIASEC Test and how it works briefly. Section 3 discusses the proposed method of soft computational technique in correlation with the RIASEC Test to determine program of study recommendations. Here, Sect. 3 is regarded as the main contribution in this paper. To clarify how each step of the proposed method's calculation process works, an illustrative example is given in Sect. 4. Section 5 proposes a flowchart to implement the development of a study program recommendation system based on a soft computational technique that has been discussed in the previous section. Finally, Sect. 6 summarizes the discussion and provides some suggestions in conclusion.

2 RIASEC

RIASEC (*Realistic*, *Investigative*, *Artistic*, *Social*, *Enterprising*, and *Conventional*) is s a psychological test tool used to determine the type of work/career and recently it is also used to determine the right program of study based on one's personality. The RIASEC test kit was first introduced by John Holland in 1970 and revised in 1977, 1985, and 1994 [1]. Based on the book SDS (Self-Directed Search): Form R 4th Edition, 1994, published by PAR (Psychological Assessment Resources), Inc., it has a validity value between 0.90 and 0.92 and a reliability value of 0.76–0.89 [2]. Based on Holland's theory [3], individual interactions with the environment can produce the characteristics of job choices and adjustments to the work environment. Holland categorized individuals into six pure personality types, namely RIASEC, where types R (Realistic), I (Investigative), A (Artistic), S (Social), E (Enterprising), and C (Conventional). The following is the definition of each personality type on the RIASEC test kit.

According to Holland's theory, someone who has a *realistic personality* is considered to prefer to work manually and enjoy operating devices (machines) or vehicles. Many realistic occupations involve outdoor activities and working with objects. Professions suitable for these realistic personalities contrast with office jobs, where people tend to sit all day indoors. According to the RIASEC theory, a person with a realistic personality tends to have hobbies such as gardening, keeping pets, mechanics, and outdoor sports.

Holland believes that people with *investigative personalities* have a strong curiosity, love to learn, and are analytical. Professions that are investigative in nature require one to analyze tasks very carefully. Hobbies of people with investigative personalities, including playing games related to training logical thinking and strategy, such as chess and sudoku games. The types of professions suitable for investigative personalities are researchers, philosophers, educators, and innovators.

Someone who gets an *artistic personality* type at the end of the RIASEC test tends to be creative and authentic. The artistic profession generally involves design, art, and expression. Various jobs as an artist require someone to see the world from a variety of new, unique perspectives, be it through their own images, language, or ideas. The hobbies of someone with an artistic personality are writing, painting, photography, crafting and design. A person with the artistic personality type can find many professions in theater, advertising, the music industry, and design firms.

If a person acquires a *social personality* after taking the Holland test, this may indicate that the person enjoys helping or teaching others. Professions that are suitable for this RIASEC personality type generally require someone to work in a teamwork. People with a social personality generally have hobbies, such as playing sports on a team, volunteering, and hanging out with friends or family. Professions suitable for social personalities can be found in government, education, health care, and social services.

People who are active, have initiative, are risk-taking, or have good leadership skills are characteristics of the *enterprising personality* type. Professions that are suitable for enterprising personalities generally require a person to make many important decisions, direct or convince others, to take responsibility for the tasks at hand. People with enterprising personalities tend to have hobbies, such as investing, training, and coaching in sports clubs, to being active in politics. Suitable jobs for entrepreneurial personalities can be found in management, marketing, commerce, administration, and politics.

People with *conventional personalities* tend to be perfectionists and prefer to work in a structured manner and according to predetermined agreements. Professions that are in accordance with conventional personalities generally require work to be carried out according to predetermined patterns or rules. This work tends to be routine, or process based. Professions that are suitable for conventional individuals can be found in the world of banking, companies engaged in property management and maintenance, administrative companies, the financial sector, to government organizations.

Table 1 Corresponding questions between RIASEC types and sections.

RIASEC	Activities	Competencies	Occupations	Self-1	Self-2
Realistic (R)	11	11	14	1	1
Investigative (I)	11	11	14	1	1
Artistic (A)	11	11	14	1	1
Social (S)	11	11	14	1	1
Enterprising (E)	11	11	14	1	1
Conventional (C)	11	11	14	1	1

Correspondence between each RIASEC type with three sections (activities, competencies, occupations) and two self-estimates ratings resulted in a total of 228 question items. Table 1 shows the details of the number of questions because of the correspondence between RIASEC types on the one side and sections and two self-estimates on the other. Thus, there are 66 activity questions (11 questions for each RIASEC type), 66 competency questions (11 questions for each RIASEC type), 84 occupation questions (14 questions for each RIASEC type), 6 self-estimates one questions, and 6 self-estimates two questions. Answers on sections, activities, competencies, and occupations use a 2-point Likert scale ("like" or "dislike" for Activities, "yes" or "no" for Competencies and Occupations), while answers on self-estimates one and two use a 7-point Likert scale (1 is low and 7 is high).

2.1 RIASEC and Program of Study

The *Self-Directed Search (SDS), Form R 4th Edition, The Occupation Finder* written by John L. Holland book [4] introduced a guide that can link RIASEC test results with the right type of occupations according to talents, interests, and personality. A pattern in the form of a code consisting of three letters called Holland Occupational Codes (HOC) is used to describe the matched occupation. Since the occupation has a relationship with the program of study to prepare student competencies to be able to work according to the job, HOC may consider corresponding to the program of study. Therefore, the second section of SDS book provided the relation between HOC and program of studies. Some examples of the relationship between the HOC and the university's 4-year (undergraduate) program can be seen in Table 2.

Table 2 Relation between HOC and program of studies.

HOC	Program of studies
CEI	Accounting
CES ERI ERS ESR	Administrative/Secretarial Science System Engineering Art History, Business Administration, Culinary Arts, Hotel Management Marketing Management, Museum Studies, Public Administration
ESI	Law, Public Finance, Political Science
ESA	Advertising, Juridical Science, Mass Communication, Travel Management
ESC CSE	Health Service Administration, Risk Management, Taxation Fiber, Textile and Weaving Arts, Legal Administration
REI RSC RIC IRE ISE ICS	Horticulture Science, Landscaping Management, Nursery Management Nuclear Technology Civil Engineering, Anthropology, Archeology, Astronomy, Civil Engineering, Neuroscience, Computer Engineering, Mechanical Engineering, Mathematics Economics, Criminology, Psychology, Medical Technology Computer Science

For example, to become an accountant, the most mandatory type of RIASEC is Conventional, followed by Enterprising and Investigative. Thus, the accounting study program as a requirement to become an accountant also has the same HOC, namely CEI. HOC can have a correlation with more than one study program. For example, ISE has a correlation with Economics, Criminology, Psychology, and Medical Technology.

On the other hand, it is possible for a study program to have more than one HOC. For example, Civil Engineering has two HOCs namely RIC and IRE. Thus, HOC and study programs have a many-to-many relationship.

3 Proposed Method of Soft Computational Technique

In contrast to the initial concept of RIASEC, answers to questions from all sections (Activities, Competencies, Occupations, Self-estimates) use a 7-point Likert scale (0 is low and 6 is high) as shown in Table 3.

Table 3 Corresponding questions between RIASEC types and sections.

Value	Activities	Competencies and occupation	Self-1 and Self-2
0	Strongly disagree	Strongly no	None
1	Disagree	No	Low
2	Somewhat disagree	Somewhat no	Slightly low
3	Neither agree nor disagree	Neither no nor yes	Neutral
4 5	Somewhat agree Agree	Somewhat yes Yes	Moderately high Very high
6	Strongly agree	Strongly yes	Extremely high

The first step is to calculate the weight of each RIASEC type. Let W_R, W_I, W_A, W_S, W_E, $W_C \in [0, 1]$ are defined as the weights of *Realistic*, *Investigative*, *Artistic*, *Social*, *Enterprising*, and *Conventional* consecutively. The weight calculation for each RIASEC type is carried out as follows. For $\alpha \in \{R, I, A, S, E, C\}$,

$$W_\alpha = \frac{W_\alpha(Ac) + W_\alpha(Co) + W_\alpha(Oc) + W_\alpha(Se1) + W_\alpha(Se2)}{30}. \tag{1}$$

Here, $W_\alpha(Ac)$, $W_\alpha(Co)$, $W_\alpha(Oc)$, $W_\alpha(Se1)$, and $W_\alpha(Se2)$ are the weights of RIASEC type α in relation to Activities, Competencies, Occupations, Self-estimates one, and Self-estimates two sections consecutively. The number 30 is the result of multiplying 5 (number of sections) and 6 (normalized so that the value is in the range 0–1). As written in Table 1, activities and competencies have 11 questions in correlation with each type of RIASEC. Therefore, the values of $W_\alpha(Ac)$ and $W_\alpha(Co)$ are obtained by calculating with the following equation.

$$W_\alpha(Ac) = \frac{\sum_{i=1}^{11} R_\alpha(Ac, i)}{11}. \tag{2}$$

Similarly,

$$W_\alpha(Co) = \frac{\sum_{i=1}^{11} R_\alpha(Co, i)}{11}. \tag{3}$$

$R_\alpha(Ac, i) \in \{0, 1, 2, 3, 4, 5, 6\}$ is the rating of the user's answer to the i-th question, the relationship between type α and activities. In the same way, $R_\alpha(Co, i) \in \{0, 1, 2, 3, 4, 5, 6\}$ is the rating of the user's answer to the i-th question, the relationship between type α and competencies. Since occupation has 14 questions for each type of α, the equation for calculating the value of $W_\alpha(Oc)$ is as follows.

$$W_\alpha(Oc) = \frac{\sum_{i=1}^{14} R_\alpha(Oc, i)}{14}. \tag{4}$$

Similarly, $R_\alpha(Co, i) \in \{0, 1, 2, 3, 4, 5, 6\}$ is the rating of the user's answer to the i-th question, the relationship between type α and occupations. As shown in Table1, Self-estimates one and Self-estimates two only have one question each. Therefore, the values of $W_\alpha(Se1)$ and $W_\alpha(Se2)$ are obtained directly from the user rating value of the relation between α with Self-estimates one, and Self-estimates two consecutively.

From the results of calculating the W_α value, the corresponding RIASEC types are sorted from the largest to the smallest. Let's say we get a RIASEC Type sequence from largest to smallest is $\alpha_1 \to \alpha_2 \to \alpha_3 \to \alpha_4 \to \alpha_5 \to \alpha_6$. If the pattern of a study program x is $\alpha_2\alpha_4\alpha_5$, then the weight of the study program x denoted by $\partial_x \in [0, 1]$ is simply calculated by the following equation.

$$\partial_x = \frac{W_{\alpha_2} + W_{\alpha_4} + W_{\alpha_5}}{3}. \tag{5}$$

It is possible for a study program to have more than one HOC, for example Civil Engineering has two HOCs, namely RIC and IRE (see Table 2). If study program x has two HOCs, namely $\alpha_2\alpha_4\alpha_5$ and $\alpha_3\alpha_4\alpha_6$, then the weight of study program x can be simply calculated as follows.

$$\partial_x = \max\left(\frac{W_{\alpha_2} + W_{\alpha_4} + W_{\alpha_5}}{3}, \frac{W_{\alpha_3} + W_{\alpha_4} + W_{\alpha_6}}{3}\right). \tag{6}$$

The greater the value of ∂_x, the better the study program x is in accordance with the talents and interests of students. Thus, the order of the study program recommendations can be given in accordance with the order of weight expressed by ∂_x. To clarify how each step of the calculation process of this method works, an illustrative example is given in Sect. 4.

4 Illustrative Example

The results of testing sections Activities and Competencies of 11 questions for each type of RIASEC are presented in Table 4. The results have been converted to values according to the answers given as shown in Table 3. For instance, the answer to the first question in the Realistic type and Activities section is "Disagree" so it is given a value of 1. The answer to the first question in the Investigative type and Competencies section is "Strongly Yes" so it is given a value of 6. Table 5 shows the test results from the Occupation section which consists of 14 questions for each RIASEC Type.

Table 4 Results of RIASEC test for activities and competencies.

Quest	Activities						Competencies					
	R	I	A	S	E	C	R	I	A	S	E	C
1	1	2	4	3	0	5	6	6	5	6	2	3
2	4	1	0	1	1	3	2	5	1	4	4	2
3	2	1	2	4	5	1	0	2	4	0	1	2
4	3	3	0	6	1	6	5	4	5	1	5	5
5	3	0	0	0	6	4	1	1	3	0	1	3
6	1	1	6	3	1	4	0	0	6	5	2	2
7	5	4	5	2	1	0	3	5	2	2	3	5
8	6	1	5	6	2	1	2	2	5	0	6	3
9	1	6	4	1	0	6	1	3	3	3	6	2
10	0	3	3	5	5	2	0	4	1	1	0	1
11	5	0	6	1	3	1	2	1	2	4	0	6

Table 5 Results of RIASEC Test for occupation.

Question	R	I	A	S	E	C
1	1	3	3	6	6	5
2	1	1	2	4	6	4
3	2	3	3	1	4	1
4	0	3	3	3	2	6
5	2	4	4	2	5	0
6	1	5	1	4	1	4
7	5	1	3	6	3	1
8	3	3	1	2	0	1
9	5	4	4	2	4	0
10	3	1	5	0	4	6
11	6	1	4	2	6	5
12	3	3	1	2	3	1
13	5	0	6	0	6	4
14	3	0	1	3	6	4

Table 6 Results of RIASEC test for self-estimates (Self-1 and Self-2).

	R	I	A	S	E	C
Self-1	5	2	1	3	4	6
Self-2	3	1	4	0	5	6

Finally, the test results for Self-estimates one and Self-estimates two are shown in Table 6. Self-estimate one and Self-estimate two have only one question related to each RIASEC Type.

By using Eqs. (2), (3) and (4), the values of $W_\alpha(Ac)$, $W_\alpha(Co)$ and $W_\alpha(Oc)$ are calculated. For $\alpha \in \{R, I, A, S, E, C\}$, value $W_\alpha(Ac)$ is calculated as follows.

$$W_R(Ac) = \frac{1+4+2+3+3+1+5+6+1+0+5}{11} = \frac{31}{11} = 2.82,$$

$$W_I(Ac) = \frac{2+1+1+3+0+1+4+1+6+3+0}{11} = \frac{22}{11} = 2.00,$$

$$W_A(Ac) = \frac{4+0+2+0+0+6+5+5+4+3+6}{11} = \frac{35}{11} = 3.18,$$

$$W_S(Ac) = \frac{3+1+4+6+0+3+2+6+1+5+1}{11} = \frac{32}{11} = 2.91,$$

$$W_E(Ac) = \frac{0+1+5+1+6+1+1+2+0+5+3}{11} = \frac{25}{11} = 2.27,$$

$$W_C(Ac) = \frac{5+3+1+6+4+4+0+1+6+2+1}{11} = \frac{33}{11} = 3.00.$$

In the same way, the results of calculations $W_\alpha(Ac)$, $W_\alpha(Co)$ and $W_\alpha(Oc)$ can be seen in Table 7.

Table 7 Calculation results of $W_\alpha(Ac)$, $W_\alpha(Co)$ and $W_\alpha(Oc)$.

	R	I	A	S	E	C
Activities (Ac)	2.82	2.00	3.18	2.91	2.27	3.00
Competencies (Co)	2.00	3.00	3.36	2.36	2.73	3.093.00
Occupations (Oc)	2.86	2.29	2.93	2.64	4.00	6.00
Self-1 ($Se1$)	5.00	2.00	1.00	3.00	4.00	6.00
Self-2 ($Se2$)	3.00	1.00	4.00	0.00	5.00	

By using Eq. (1), the weights of *Realistic*, *Investigative*, *Artistic*, *Social*, *Enterprising*, and *Conventional* can be calculated as follows.

$$W_R = \frac{2.82+2.00+2.86+5.00+3.00}{30} = 0.52, \quad W_I = \frac{2.00+3.00+2.29+2.00+1.00}{30} = 0.34,$$
$$W_A = \frac{3.18+3.36+2.93+1.00+4.00}{30} = 0.48, \quad W_S = \frac{2.91+2.36+2.64+3.00+0.00}{30} = 0.36,$$
$$W_E = \frac{2.27+2.73+4.00+4.00+5.00}{30} = 0.60, \quad W_C = \frac{3.00+3.09+3.00+6.00+6.00}{30} = 0.70.$$

From the calculation results above, the RIASEC Type order based on the largest weight is as follows: $C \to E \to R \to A \to S \to I$. It can be verified that some of the HOCs in Table 2 comply with the order of the survey results are CEI, CES, ERI, ERS, and ESI. By using Eq. (5), the weight of each study program x can be calculated based on its HOC as follows.

- $\partial_x = \frac{0.7+0.6+0.34}{3} = 0.546$, for all x satisfy CEI.
- $\partial_x = \frac{0.7+0.6+0.36}{3} = 0.553$, for all x satisfy CES.
- $\partial_x = \frac{0.6+0.52+0.34}{3} = 0.486$, for all x satisfy ERI.
- $\partial_x = \frac{0.6+0.52+0.36}{3} = 0.493$, for all x satisfy ERS.
- $\partial_x = \frac{0.6+0.36+0.34}{3} = 0.430$, for all x satisfy ESI.

The order of the largest HOC weight that meets the survey results is as follows: $CES \rightarrow CEI \rightarrow ERS \rightarrow ERI \rightarrow ESI$. Finally, study program recommendations based on survey results can be seen in Table 8.

Table 8 Recommendation of study programs based on the survey.

Rec	HOC	Program of studies
#1	CES	Administrative/Secretarial Science
#2	CEI	Accounting
#3	ERS	Art History, Business Administration, Culinary Arts, Hotel Management
#4	ERI	System Engineering
#5	ESI	Law, Public Finance, Political Science

5 Flowchart of System Recommendation

This section describes the system design in a flowchart to develop a recommendation system. First, prospective students can log in using their personal email and password. On the main page, users can see several menu options. The first menu is the User Profile Menu. Users can register by entering and changing personal data on the User Profile menu. The second menu is the Self Interest and Personality Test Menu. In this menu, the user can carry out the interest and personality test process by selecting one of the seven answer choices that have been provided.

At the test stage, the user will be faced with different questions for each of the six types of questions (RIASEC) which are classified into five sections (Activity, Competence, Occupation, Self-1, and Self-2). After carrying out the test, the Interest Test menu will display the results of study program recommendations based on prioritized interests using soft computational techniques as discussed in the previous section.

The results of Study Program recommendations can be downloaded by users in the form of pdf files and can be sent via the user's personal email. The last menu is the Contact Us Menu. This menu only provides the address of the university and contact number and place of consultation if needed.

6 Conclusion

This paper discussed and introduced a soft computing technique to build a study program selection recommendation system based on the RIASEC Test. In contrast to the original proposed RIASEC Test which uses a 2-point Likert scale, the RIASEC Test used in this paper uses a 7-point Likert scale to obtain softer and more accurate survey results. Several simple formulas were introduced to evaluate survey results and calculate the weight of study programs used to express suitability for interests and talents. The suitability weight of the study program is used to determine the level of recommendation. An illustrative Example was given and discussed to facilitate understanding of the proposed concepts. A computer-based recommendation system has been built and implemented according to the existence of study programs at Petra Christian University and has received a very positive response from prospective students. To produce a more comprehensive and reliable recommendation system, the recommendation results based on the RIASEC Test need to be integrated with the student's academic qualifications.

References

1. Dozier, V.C., Sampson, J.P., Reardon, R.C.: Using two different self-directed search (SDS) interpretive materials: implications for career assessment. Professional Counselor **3**(2), 67–72 (2013)
2. Dozier, V.C., Sampson, J.P., Lenz, J.G., Peterson G.W., Reardon, R.C.: The impact of the self-directed search form R internet version on counselor-free career exploration. J. Career Assess. 1–15 (SAGE, Florida State) (2014)
3. Holland, J.L.: Making Vocational Choices: A Theory of Vocational Personalities and Work Environments, 3rd edn. Psychological Assessment Resources, Odessa, FL (1997)
4. Holland, J.L.: The occupations finder. Psychological Assessment Resources, Odessa, FL (2000)
5. Armstrong, P.I., Day, S.X., McVay, J.P., Rounds, J.: Holland's RIASEC model as an integrative framework for individual differences. J. Couns. Psychol. **5**(1), 1–18 (2008)

Semiconductor Manufacturing Final Test Yield Prediction Using Regression with Genetic Algorithm-Based Feature Selection

Ranzel V. Dimaculangan(✉), Robert G. de Luna, Marife A. Rosales, Jo-Ann V. Magsumbol, and Orland D. Tubola

Polytechnic University of the Philippines, M.H. del Pilar Campus, Valencia St., Sta. Mesa, Manila, Philippines
dimaculanganranzel@gmail.com, {rgdeluna,marosales,javmagsumbol, odtubola}@pup.edu.ph

Abstract. From raw silicon to a finished integrated circuit (IC), a typical semiconductor manufacturing process involves wafer fabrication, wafer probe, assembly, and final test. Before an IC is delivered to customers, it undergoes final test where its overall functionality is verified. This stage is where low yield issue is often encountered and with yield having a direct impact on a company's revenue, it is considered a critical key process indicator. To the best of the author's research, this is the first study to use both continuous and categorical front-end (excluding WAT parameters) and back-end variables to predict final test yield through regression analysis. A three-year amount of real production data was used to train, validate, and evaluate 11 different regressors. A genetic algorithm is also implemented for feature reduction. Results show that the optimized Bagging Regression model has a R2 score of 0.6813 with an average mean absolute error of 0.3902%.

Keywords: Semiconductor manufacturing · Final test yield prediction · Machine learning · Genetic algorithm · Regression

1 Introduction

Integrated circuits (ICs) are microelectronic circuits made with several types of electrical components embedded in a semiconductor. This has revolutionized the world of technology by enabling miniaturization and increasing efficiency in electronic devices. To produce an IC, a typical semiconductor manufacturing process involves wafer fabrication, wafer probe, assembly, and final test.

Wafer fabrication is the process of creating ICs on a thin semiconductor material called a wafer. It involves a series of steps such as deposition, photolithography, etching, and doping. The result is a wafer with many identical ICs that can be separated into individual chips or oftentimes referred to as a die. Also during this stage, a Wafer Acceptance Test (WAT) is conducted to monitor important process parameters related to wafer fabrication, and only wafers passing through it will proceed to wafer probe [1].

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 108–117, 2024.
https://doi.org/10.1007/978-3-031-50151-7_11

Meanwhile, wafer probe is a test method used in the semiconductor industry to evaluate the electrical performance of ICs on a wafer before they are packaged and sold. The wafer is covered with electrical probes that make contact with the individual ICs on the wafer. The probes are connected to a test equipment that applies signals to the ICs and measures the response to determine if the device meets specifications.

The third stage of the semiconductor fabrication chain refers to the process of packaging dies which is called assembly. This process involves attaching the ICs to a package that provides protection and interconnects for the device. The package can be as simple as a metal can, plastic, or ceramic compound.

Lastly, the packaged dies will go through final test which is the process of thoroughly testing a product to ensure that it meets all specifications and requirements before it is shipped to customers. In the semiconductor industry, final test involves applying a set of tests to each packaged device to verify its electrical performance, functionality, and reliability. The tests can be automated and performed using specialized test equipment. The results of the final test are used to determine if the device meets quality standards and can be shipped to customers.

Nowadays, this already complex semiconductor manufacturing continues to get more challenging with the shrinking technological nodes and process improvements. Along with the increased complexity is the rising cost which was also brought by global supply shortages; Hence, making manufacturing yield optimization one of the most important objectives for semiconductor fabrication [2]. With final test as the last layer of verification done on a product, predicting it at an earlier stage will increase productivity and reduce scrap-induced shortages which could result in delayed or failed deliverables to customers.

2 Review of Related Literature

From the front-end process (wafer fabrication and probe) to the back-end (assembly and final test), several potential causes can result in a low final test yield. In [1], WAT parameters and categorical data such as wafer technology, RAM/ROM versions, firmware versions, package types, product functionality, fab and test locations, test program versions, and tester and handler types are used to classify final test yield which was clustered through Gaussian Mixture Model (GMM). One improvement that can be taken from this study is to predict final test yield as a numerical value instead of bins with different ranges as it can give more valuable information to potential output quantities at the backend.

Meanwhile, only WAT parameters are used in [3] to predict final test yield as a continuous parameter by training and evaluating different regression models. From a 10-fold cross-validation, such study attained the best average root-mean-square error (RMSE) of 4.005 using XGBoost Regressor. As stated in [3], categorical information such as product configuration, fabrication factors, and test equipment can influence the final test yield. Thus, the main motivation of this study is to fill in such research gap by using both front-end (excluding WAT parameters) and back-end numerical and categorical data to predict final test yield as a numerical value through regression.

With this study, the other sources of process variations introduced during semiconductor manufacturing will be taken into account in evaluating their correlation with final test yield. Utilizing such a model that is capable of predicting final test yield at an earlier stage of the process, can benefit semiconductor manufacturing organizations to have a more effective planning and forecasting schedule. An ability to predict final test yield means an ability to predict shippable output quantities to meet customer deliverables. This will allow more efficient allocation of resources and equipment which leads to timely fulfillment of orders.

Detecting low-yielding backend lots before assembly will also save packaging costs. Scrapping rejects lots ahead of time will also save production resources and will boost throughput; Thus, less operational costs and increased productivity translate to an increased profit margin [4]. Moreover, yield improvement activities can also be taken proactively. Identifying lots with low final test yield before the process itself can pre-alert engineering teams and then take preventive measures to determine the potential causes of the issue and eventually resolve it if feasible.

3 Methodology

The raw dataset to be used in this study is provided by Microchip Technology Inc. It is comprised of 392,648 backend lots from 10,362 different 8-bit PIC microcontroller products that underwent final test from 2020 to 2022.

3.1 Data Wrangling and Exploratory Data Analysis

One of the major parts of the data wrangling process is to identify and handle missing values. Rows with null values that cannot be imputed or derived from existing features will be dropped such as the target variable as it does not add any meaningful information to the dataset. On the other hand, missing values will be imputed when the substitution is related to other variables and deemed as an acceptable method for estimation such as using the mean or mode of the non-missing values.

Meanwhile, an Exploratory Data Analysis (EDA) will be done to analyze and summarize the dataset to gain insights, identify patterns, and detect anomalies [5]. Through this process, features can already be reduced by calculating their statistical significance to the target variable.

For numerical features, the Pearson correlation coefficient will be calculated [6]. This value measures the linear dependence between the independent and dependent variables. The resulting coefficient is a value between -1 and 1 which indicates a direction from a total negative to a total positive linear correlation.

For categorical features, the F-test score will be computed. This value is used to evaluate if there is a significant difference between the means of two or more groups. By using analysis of variance (ANOVA), the F-test metric will provide a measure of how well the model fits the data, where a high F-test score indicates that the means of the groups are significantly different.

The statistical significance of the Pearson correlation coefficients of the numerical variables and the F-test score of the categorical variables will be used to eliminate features [7]. In this study, a predictor variable will be accepted if it exceeds the threshold set by a significance level of 99%. It means that such scores are statistically significant if their p-value is less than 0.01.

All categorical variables will be converted into continuous parameters using One-Hot Encoding [1]. It is considered as an unsupervised encoding method since the features will rely on their corresponding values alone [8]. Using all features, different regression models will then be evaluated based on regression performance metrics. Only the top three regressors will be used in the feature selection stage.

3.2 Feature Selection

A genetic algorithm approach will be used to reduce the number of predictor variables. By identifying a subset of features that will best improve the performance of the machine learning model, overfitting and the computation complexity of the model will be reduced [9]. The basic flow of this process is shown in Fig. 1.

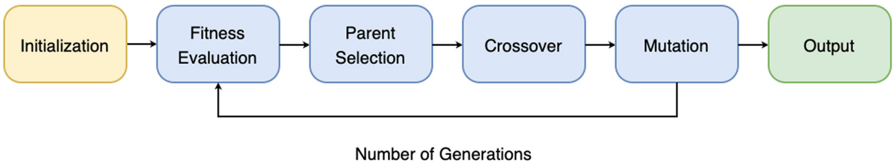


Fig. 1 Genetic algorithm basic process flow

On initialization, the first set of the population is randomly generated with a fixed size, each representing a candidate solution (also called individuals or genome). The genome represents a subset of features while each gene represents a binary value indicating whether the corresponding feature is included or excluded in the subset [10]. Thus, the genome width is equal to the number of features.

Each genome will be evaluated using a fitness function that measures the average R2 score using cross-validation. Afterward, the parents will be selected as the top two individuals with the best score. They may form offspring through a genetic operator of crossover where the genes of the parents are combined. The child can also mutate which introduces random variations by flipping a gene from a binary value of 0 to 1 or vice versa. The processes between fitness evaluation to mutation are repeated depending on the number of set generations [11].

The genetic algorithm will output the best individual from the last generation with the highest fitness value (average R2 score) as the final solution. This procedure will be conducted thrice using each of the three selected regressors.

3.3 Model Optimization and Evaluation

Hyperparameters vary from one algorithm to another and are set before training a model [12]. Using a brute force method, all possible combinations of the hyperparameters of the machine learning model will be assessed based on the average R2 score of cross-validation. Afterward, the optimized regressors will be finally evaluated using other performance metrics which include mean absolute error (MAE), mean squared error (MSE), and explained variance score (EVS).

MAE measures the average absolute difference between the predicted and true values, while MSE measures the average of the squared value of such differences. Moreover, root-mean-square error (RMSE) is the square root value of MSE which penalizes larger errors, making it sensitive to outliers. Meanwhile, EVS measures the proportion of variability in the predictions explained by the independent variables as it is computed as the ratio of the difference between the actual and predicted value to the total variance in the actual values.

4 Results

Following the methodology in Sect. 3, the results are divided into three stages: (113.1) Data Wrangling and Exploratory Data Analysis, (113.2) Feature Selection, and (113.3) Model Optimization and Final Evaluation.

4.1 Data Wrangling and Exploratory Data Analysis

With five other datasets from Microchip Technology Inc., a total of 20 front-end and back-end parameters were extracted and selected as potential predictors. Fourteen of these are categorical variables which include the Mask, Package, Restriction Level, Process Plan, Test Method, Assembly Site, Wafer Fabrication Site, Product Grade, Tester Type, Pin Count, Number of QC Steps, Insertion Count, Site Count, and Part Number. The remaining six are continuous parameters.

Upon consolidating the dataset, a total of 334,837 missing values were found in the columns of Test Method, Average Probe Yield at Hot and Room Temperature, Run Rate, Part Number, and Process. Due to no possible imputation method, all rows that contain null values were dropped except for the Test Method. Missing values in the said column are derived from the existing Process Plan feature.

Due to infrastructure constraints, the training data was also trimmed down to only those masks with 2500 data points or more to minimize the run time for feature selection and machine learning modeling. Thus, the number of lots was reduced to 42,866.

Afterward, the F-test scores of all 14 categorical variables were calculated. Insertion Count, Test Method, and Package have the highest F-test scores of 1503.2502, 1440.6242, and 11,104.7449, respectively. It indicates that the average yield in each level of these features is significantly different from one another.

On the other hand, Tester Type has the lowest F-test score of 2.6267. It is the only feature with a high p-value of 0.1 that fails to meet the p-value criterion of less than 0.01 for a confidence interval of 99%. However, it will still be retained and further evaluated through the genetic algorithm-based feature selection method.

Meanwhile, it was observed that there is a strong correlation of 0.78 between the Average Probe Yield at Hot Temperature and the Average Probe Yield at Room Temperature. This means that the majority of wafer lots with a good yield at room temperature also yield satisfactorily at hot temperature. Moreover, a strong correlation of 0.96 is also expected between Die Width and Die Length as the die shape is typically rectangular or square due to design and other engineering considerations.

Table 1, on the other hand, indicates that all six numerical features have a weak correlation with final test yield, but are all significant as the *p*-value is zero. Thus, all continuous parameters will be used in the machine learning model development.

Table 1 Summary of correlation scores and *p*-values of numerical variables

Feature	Corr. score	*p*-value
Input quantity	0.1501	0.0000
Ave. probe yield at hot temp	0.1207	0.0000
Ave. probe yield at room temp	0.1139	0.0000
Die width	0.2162	0.0000
Die length	0.1794	0.0000
Run rsate	0.2598	0.0000

Afterward, the 14 categorical variables are converted into continuous parameters using the One-Hot Encoding technique. The number of independent variables increased from 20 to 184. One column from each of the encoded features was removed to prevent a dummy trap, thus decreasing the number of columns to 170.

Using all unscaled features in a five-fold shuffled cross-validation, the R2 scores of the 11 different regression models were calculated. A negative R2 score was recorded using Support Vector and AdaBoost Regressors which implies that the model is not a good fit for the data and its predictions are far from true values.

As summarized in Table 2, the effect of feature scaling is then tested by computing the average R2 score and mean absolute error (MAE) of the regression models before and after feature scaling. Feature scaling reduced the performance of most of the machine learning models as it amplified the effects of outliers in the dataset. However, the outliers cannot be removed as it means removing backend lots with a low yield which will decrease the reliability of the model to predict low-yielding lots. On contrary, feature scaling greatly improved the R2 score of the K Neighbors Regressor from 0.3767 to 0.5422 and the SVR from − 0.0007 to 0.4308. Thus, the top three models are Random Forest Regressor, Bagging Regressor, and K-Nearest Neighbor Regressor where feature scaling will only be applied to the first two algorithms.

4.2 Feature Selection

Using all 170 features, the top three regressors in their default state are fed into a feature selection process using a genetic algorithm. It was run through 30 generations with an initial population size of 15, one-point crossover with a probability of 80%, and one-gene bit flip mutation with a probability of 20%. The fitness function evaluates an individual based on its average R2 score from a five-fold shuffled cross-validation.

As shown in Fig. 2, the R2 score optimization using Random Forest seems to not reach convergence yet after 30 generations and only increased the metric by 0.31% from 0.6520 to 0.6540. Conversely, the right graph shows a plateau in the fitness function using

Table 2 R2 score and MAE comparison before and after feature scaling

Model	Before		After	
	R2	MAE	R2	MAE
Random forest	**0.6547**	0.3864	**0.6541**	**0.3865**
Bagging	**0.6187**	0.4035	**0.6185**	**0.4037**
Gradient boosting	**0.4976**	0.5261	0.4976	0.5261
Decision tree	0.4599	0.4663	0.4581	0.4667
Ridge	0.4345	0.5473	0.4345	0.5472
Linear	0.4344	0.5473	− 1.74E+20	2.32E+08
K neighbors	0.3767	0.5967	**0.5422**	**0.4537**
Elastic net	0.0826	0.8194	− 0.0001	0.8466
Lasso	0.0797	0.8207	− 0.0001	0.8466
SVR	− 0.0007	0.7272	0.4308	0.4840
Ada boost	− 1.5323	1.8076	− 1.5857	1.8360

The highlighted numbers are the top three scores in each performance metric

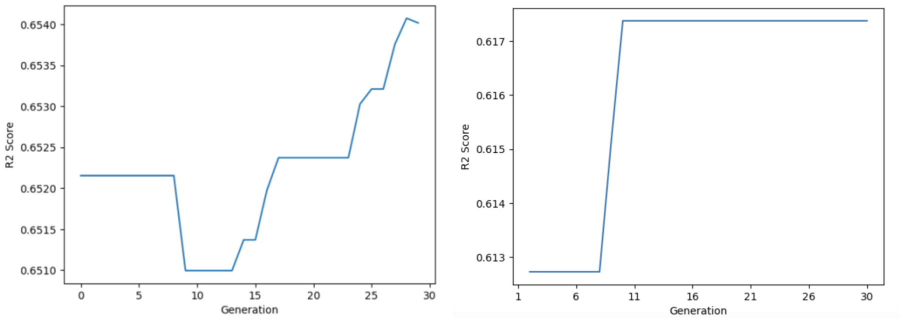


Fig. 2 R2 score over generations using random forest (*Left*) and Bagging regressor (*Right*)

Bagging Regressor which implies that it has reached convergence. For 20 consecutive generations, the average R2 score using Bagging Regressor has stagnated at 0.6174 but it is still an improvement of 0.63% from 0.6135.

Similarly, the genetic algorithm using K Neighbors Regressor is no longer making progress after the 16th generation as shown in Fig. 3. The fitness value of 0.5564 remained unchanged for the next 14 generations. Nevertheless, its average R2 score increased by 0.14% from 0.5556.

Along with the minimal improvement in the model's performance due to feature selection, the number of predictor variables was also trimmed down as shown in Table 3. Meanwhile, either the Die Width or Die Length was dropped using the three regressors as expected due to their strong correlation. Also, only the Random Forest Regressor dropped the Tester Type feature with a low F-test score and high *p*-value. Table 3 also

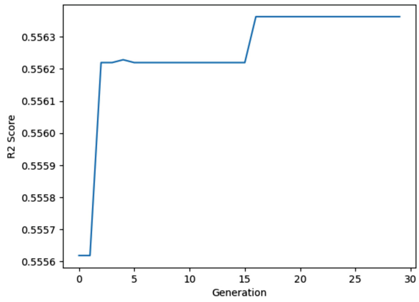


Fig. 3 R2 Score over generations using K neighbors regressor

shows the best R2 score, the number of crossover and mutations, and the time lapsed to complete the algorithm using the three selected regressors.

Table 3 Results of genetic algorithm-based feature selection

Model	Retained	Categorical	R2 score	Crossover	Mutation	Run time
RF	95	90	0.6540	24	15	3.34 h
Bag	96	92	0.6174	28	19	0.40 h
KNN	97	93	0.5564	27	19	1.44 h

The selected features of each regression model were saved into different arrays and were further evaluated as summarized in Table 4.

Table 4 Average and standard deviation of performance metrics using the selected features

Metric	10-fold cross validation					
	Average			Standard deviation		
	RF	Bag	KNN	RF	Bag	KNN
R2	**0.6540**	0.6174	0.5564	0.0287	0.0258	**0.0246**
MAE	**0.3885**	0.4107	0.4566	0.0078	0.0103	**0.0044**
MSE	**0.7639**	0.8443	0.9783	0.0818	0.0762	**0.0742**
EVS	**0.6542**	0.6175	0.5564	0.0286	0.0257	**0.0246**

Unanimously, the Random Forest yielded the highest average value while the K Neighbors Regressor produced the lowest standard deviation among all the four performance metrics.

4.3 Model Optimization and Evaluation

For hyperparameter tuning, a brute-force approach using GridSearchCV was implemented to find the optimal settings of the three selected regressors. Using the set of optimized hyperparameters, the average R2 score of each regression model after a five-fold shuffled cross-validation was calculated. Results show that Bagging Regressor has the best fit of the model to the data using the following optimized hyperparameters: Number of estimators = 100, Bootstrap = True, Bootstrap features = True, Maximum features = 1, Maximum samples = 1, and Warm start = True.

To further evaluate the optimized regressors, Table 5 summarizes five performance metrics using 10-fold shuffled cross- and hold-out validation. It can be observed that the Bagging Regressor garnered the highest scores in all metrics except the MAE.

Table 5 Performance metrics using the optimized regressors on selected features

Metric	10-fold cross validation						Hold-out validation		
	Average			Standard deviation					
	RF	Bag	KNN	RF	Bag	KNN	RF	Bag	KNN
R2	0.6725	**0.6813**	0.6153	0.0418	**0.0335**	0.0534	0.6084	**0.6363**	0.5443
MAE	**0.3794**	0.3902	0.4079	**0.0073**	0.0088	0.0131	**0.3799**	0.3875	0.4089
MSE	0.7259	**0.7058**	0.8486	0.1341	**0.1175**	0.1377	0.8964	**0.8325**	1.0431
EVS	0.6726	**0.6815**	0.6154	0.0419	**0.0336**	0.0533	0.6084	**0.6363**	0.5444
RMSE	0.8485	**0.8372**	0.9182	0.0768	**0.0689**	0.0741	0.9468	**0.9124**	1.0213

5 Conclusions

This research was able to evaluate the effectivity of using both front-end (excluding WAT parameters) and back-end numerical and categorical data to predict final test yield as a numerical value through regression modeling. After evaluating 11 different regression models using 170 encoded categorical features and numerical parameters, Random Forest, Bagging, and K Neighbors ranked as the top three models with R2 scores of 0.6547, 0.6187, and 0.5422, respectively, and MAE of 0.3864, 0.4035, and 0.4537, accordingly.

Through genetic algorithm-based feature selection, the dimensionality of the model using Random Forest, Bagging, and K Neighbors Regressors was reduced to 95, 96, and 97, respectively, with no significant improvement in the goodness-of-fit value. The Random Forest Regressor produced the best R2 score of 0.6540 and an MAE of 0.3885.

The three models were then optimized where the Bagging Regressor yielded the highest R2 score of 0.6813 with an MAE of 0.3902. Its RMSE of 0.8372 is also better than the RMSE of 4.005 in [3] where only WAT parameters are used as predictor variables.

From using all 170 features with the default Bagging Regressor hyperparameters to using the selected 96 features with the optimized model, the average R2 score from a cross-validation improved by 10.12% from 0.6187 to 0.6813.

References

1. Jiang, D., Lin, W., Raghavan, N.: A novel framework for semiconductor manufacturing final test yield classification using machine learning techniques. IEEE Access **8**, 197885–197895 (2020). https://doi.org/10.1109/ACCESS.2020.3034680
2. Jiang, D., Lin, W., Raghavan, N.: Semiconductor manufacturing final test yield optimization and wafer acceptance test parameter inverse design using multi-objective optimization algorithms. IEEE Access **9**, 137655–137666 (2021). https://doi.org/10.1109/ACCESS.2021.3117576
3. Jiang, D., Lin, W., Raghavan, N.: A Gaussian mixture model clustering ensemble regressor for semiconductor manufacturing final test yield prediction. IEEE Access **9**, 22253–22263 (2021). https://doi.org/10.1109/ACCESS.2021.3055433
4. Senoner, J., Netland, T., Feuerriegel, S.: Using explainable artificial intelligence to improve process quality: evidence from semiconductor manufacturing. Manage Sci **68**(8), 5704–5723 (2022). https://doi.org/10.1287/mnsc.2021.4190
5. Rawat, S., Rawat, A., Kumar, D., Sabitha, A.S.: Application of machine learning and data visualization techniques for decision support in the insurance sector. Int. J. Inf. Manag. Data Insights **1**(2), 100012 (2021). https://doi.org/10.1016/j.jjimei.2021.100012
6. Halibas, A.S., Cherian Matthew, A., Pillai, I.G., Harold Reazol, J., Delvo, E.G., Bonachita Reazol, L.: Determining the intervening effects of exploratory data analysis and feature engineering in telecoms customer Churn modelling. In: 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC) (2019). https://doi.org/10.1109/icbdsc.2019.864557
7. Sai, B. N. K., & Sasikala, T.: Predictive analysis and modeling of customer Churn in telecom using machine learning technique. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (2019). https://doi.org/10.1109/icoei.2019.886262
8. Gnat, S.: Impact of categorical variables encoding on property mass valuation. Procedia Comput. Sci. **192**, 3542–3550 (2021). https://doi.org/10.1016/j.procs.2021.09.127
9. Siddiqi, M.A., Pak, W.: Optimizing filter-based feature selection method flow for intrusion detection system. Electronics (Switzerland) **9**(12), 1–18 (2020). https://doi.org/10.3390/electronics9122114
10. Rostami, M., Berahmand, K., Forouzandeh, S.: A novel community detection based genetic algorithm for feature selection. J Big Data **8**(1), 1–27 (2021). https://doi.org/10.1186/s40537-020-00398-3
11. Sayed, S., Nassef, M., Badr, A., Farag, I.: A nested genetic algorithm for feature selection in high-dimensional cancer microarray datasets. Expert Syst Appl **121**, 233–243 (2019). https://doi.org/10.1016/j.eswa.2018.12.022
12. Pudjihartono, N., Fadason, T., Kempa-Liehr, A.W., O'Sullivan, J.M.: A review of feature selection methods for machine learning-based disease risk prediction. Front. Bioinf. **2** (June 2022). https://doi.org/10.3389/fbinf.2022.927312

Modeling and Forecasting Bank Stock Prices: GARCH and ARIMA Approaches

Leeto Malose Patrick Shogole, Sharon Nwanamidwa, Martin Chanza, Elias Munapo(✉), and Kolentino N. Mpeta

Department of Business Statistics and Operations Research, North-West University, Potchefstroom, South Africa
Elias.Munapo@nwu.ac.za

Abstract. This study aimed at modeling and forecasting volatility of the Standard Bank Stock Prices using GARCH and ARIMA models. ARIMA is the Autoregressive Integrated Moving Average and GARCH represents Generalized AutoRegressive Conditional Heteroskedasticity. The time series data used in the study is from February 2006 to July 2022 and comprises of 198 observations downloaded from *Yahoo* finance. The best model was selected using the Akaike Information Criterion (AIC) procedure. The model with the smallest AIC was selected as the best model. GARCH (1,1)-ARMA (2,2) was the best model for modeling and forecasting volatility in the Standard Bank stock prices. The selected model were used to forecast the Standard Bank monthly closing stock prices. The results obtained revealed a gradual increase in Standard Bank closing stock prices in the next year.

Keywords: ARIMA · GARCH · Stock price · Forecasting · Volatility

1 Introduction

Volatility modeling and forecasting are of utmost importance in the financial markets and have been an area of interest for researchers in recent decades. In any country's economic growth and development process, financial markets play a crucial role [12]. Financial time series associated with financial markets are noisy by nature and non-stationary, making them one of the most challenging applications. Reference [18] defines volatility of prices as the rate at which prices change. After investing in an asset, economists and other business analysts define this rate as the standard deviation from the return. It can also be defined as a proportion of changes in the dispersion of earnings or returns for a particular marketable security. Volatility can weaken the smooth operation of the financial system and has a negative impact on economic performance.

Risk is referred to as volatility, according to [15]. Most people criticize high volatility because it suggests that securities are unstable and that equity markets are not operating as efficiently as they could. A rise in stock market volatility can cause a shift to less risky funds, which is the same as an increase in the risk of stock market investments [12]. Similarly, equity market volatility also has a number of detrimental effects, such

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 118–132, 2024.
https://doi.org/10.1007/978-3-031-50151-7_12

as those on consumer spending [10]. Consumer spending will decline due to a declining stock market and decreased consumer confidence [20].

In contemporary "finance, volatility modeling and prediction is also a significant risk indicator and is essential for many financial applications, including options pricing, hedging techniques, portfolio management, and derivatives [11]. Good projections are essential for improving financial decisions due to the rising volatility of the financial markets and the globalization of capital flows. Together with traders, this will assist risk managers in minimizing risk and helping them make informed decisions. So much in terms of research has been done to improve the performance of volatility models because better forecasting leads to" improved pricing, asset management, and financial risk management [16].

The usage of GARCH models to analyze financial data has significantly improved and is now more popular with business analysts in recent years since, in the end, they correctly explain nonlinearities. According to Nguyen et al. (2019) Mean and variance are jointly modeled using Autoregressive Conditional Heteroskedasticity (ARCH). Furthermore, conditional variance is allowed by the GARCH model. In the past, high volatility problems in forecasting financial time series have been addressed using GARCH models [9]. When modeling economic and financial data, these two approaches seem to make the results more predictable. These models seem to make the modeling of economic and financial data more predictable. High-frequency data with extreme kurtosis and asymmetric effects, fluctuating volatility across time, and concentration of volatility can all be explained using GARCH models" [2]. In other" research, such as that by Vejendla and Enke [22], it was discovered that GARCH models provide an accurate forecast of volatility than neural networks.

Several models have been employed in numerous research to "forecast financial data. The Standard Bank monthly closing stock price are predicted in this study using the GARCH models. The findings of this study will enable banks to control their operations. In addition, the outcomes will enable banks to keep an eye on their revenue streams while continuing to serve clients and enterprises with high levels of debt. This study is useful to those in the banking sectors, government" planners, economists, etc.

In forecasting financial data, most of the research done so far have employed a variety of models. The monthly closing prices of the standard bank stock prices are predicted in this study using the GARCH and ARIMA models. As a result of developments in forecasting techniques, financial managers are now able to make informed and accurate decisions on their business operations. Additionally, banks will use the accurate forecasting results to monitor their revenue streams and at the same time continue offering excellent service to their customers. This study contributes to the pool of available knowledge by modeling the latest financial data and drawing conclusions which is very important for investors.

The next part of the paper is divided into four main sections. The literature review forms Sect. 2 of the paper and Sect. 3 is the research methodology. In Sect. 4 discussions of the results are presented and in Sect. 5 the study is concluded.

2 Literature Review

Reference [17] used a number of asymmetric and symmetric GARCH models to explore the volatility of the stock market. What emerged from the study is that there is a substantial positive association between risk and returns in symmetric GARCH models. There is no leverage in the rate of return, according to the asymmetric GARCH models, which also demonstrate that the estimates are significant and that the estimate for leverage is both negative and significant. The results also demonstrated that future negative shocks are less volatile than positive shocks. For the Saudi Arabian stock market, it was shown that symmetric and asymmetric GARCH models are preferable for reflecting volatility if intraday data is used.

Reference [8] used GARCH to predict daily return volatility in the Kenyan stock market via the Nairobi Securities Exchange (NSE). The leverage effect and volatility clustering were two common traits of stock markets that were studied using both symmetric and asymmetric models. The findings revealed a high consistency of volatility, proving that the returns on the NSE index include a risk premium. This lends credence to the idea that volatility and anticipated stock returns are positively correlated. In addition, the outcomes showed that asymmetric GARCH models fit NSE better than symmetric models.

Reference [13] examined the effect of a straightforward GARCH model. Using four Bangladeshi companies that are listed on the Dhaka Stock Exchange (DSE), the study used a distinct lag order of the GARCH model to simulate the volatility of stock returns. The AIC and BIC methods were applied to choose the optimal GARCH (p, q) model. The results demonstrated that the daily return distribution is non-normal, negatively skewed, and has a significant kurtosis. The results also showed that, when modeling volatility of the daily return series of the DSE, GARCH (1,1)" is the preferable and more effective model than the other GARCH (p, q) models.

Reference [1] utilized GARCH models to evaluate and predict stock market volatility from Sudan. The findings demonstrate "that there is conditional variance, while constant for the index range of the CASE series, explodes for the return series of the KSE index. A validation of the existence of a positive risk premium in both markets, which backs up the idea that volatility and predicted stock returns are positively correlated. The study also revealed that the asymmetrical GARCH offers compelling evidence of the asymmetry of stock returns in the two markets, supporting" the presence of the leverage effect in the return series.

A study was done by [3], which used GARCH (1,1) and GJR—GARCH (1,1) considering a generalized error distribution, to see if a recurrent neural network gives more precise Value at Risk "(VaR) and Expected Shortfall (ES) forecasts of the EUR/USD exchange rate than the traditional GARCH (1.1) model, based on the Basel Committee on Banking Supervision's recommendation to move from Value at Risk (VaR) to Expected Shortfall (ES) when determining Market Risk Capital. The results showed that the VaR 95% performance of the Gated Recurrent Unit neural network is superior to that of the GARCH (1,1) model, while the accuracy of the predictions made by the Gated Recurrent Unit neural network" is higher.

Reference [7] used the GARCH (1,1) and Glosten-Jagannathan-Runkle GARCH (1,1) also known as GJR-GARCH (1,1) models, considering a generalized error distribution, to analyze the "behavior of return volatility on the Nigerian Stock Exchange (GED). The overall study's findings support Nigeria's stock performance data, which show the effects of resilience, fat tail distribution, and higher profitability. Results from the $GARCH(1, 1)$ model showed that there was volatility clustering in the NSE yield chain. The $GJR-GARCH(1, 1)$ model findings additionally" demonstrated the existence of leverage effects in the series.

Reference [21] conducted research to forecast volatility using the index return of the American Stock Exchange (ASE). The study "examined a variety of forecasting techniques, including the relatively simple GARCH (1,1) model and more complicated GARCH models such as the Exponential GARCH (1,1) and Threshold GARCH (1,1). The results indicated that, despite the existence of a leverage impact, the symmetric GARCH model performs better in predicting the conditional variation of the performance of the S&P 500 index than asymmetric GRH models based on out-of-sample predictions and" the majority of assessment metrics.

Using data from the stock markets of Hong Kong, Japan, Malaysia and Indonesia, [5], examined the effectiveness of the GARCH, ARIMA and EWMA models. The findings from the study supported the idea that volatility in emerging economies is not always greater than that in developed economies. In addition, for the Hong Kong stock market, it was shown that the best forecasting is the EWMA model. On the other hand, the best forecasting model for stock markets in Malaysia, Indonesia, and Japan is the GARCH (1,1) model.

Reference [14] used the conventional GARCH (1,1) model to simulate the volatility in the Indian equities market. In order to determine if the volatility on the NSE increased following the establishment of the Volatility Index, the analysis also included a GARCH (1,1) model where a dummy was used (India VIX). For this study, the pre-IVIX introduction period was from 1st January 2000 to 31st October 2007. Then the post-IVIX introduction era was from 1st November 2007 to 31st August 2016. What emerged analysis of the GARCH (1,1) model that had a dummy is that the spot market's volatility decreased after the introduction of IVIX India. Furthermore, the results of standard GARCH (1,1) model showed that in the post-IVIX implementation era, recent news had a greater effect on spot market shift.

3 Research Methodology

In this section, the research methodology used in the study is explained. This includes the models ARCH, GARCH and ARIMA. In addition, the model selection criteria such as the AIC and ML methods of parameter estimation are described.

3.1 Research Methodology as a Flow Chart

Figure 1, shows the research methodology.

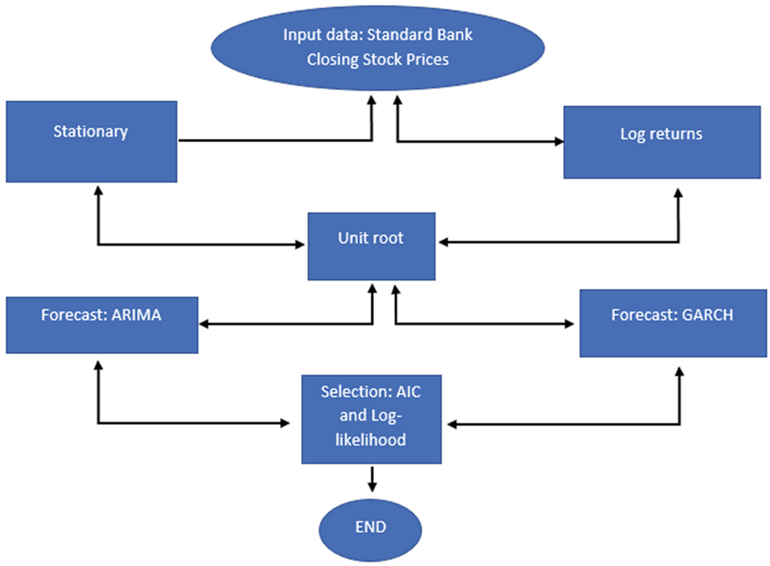


Fig. 1 Methodology flow chart [6]

3.2 Data Source

The data used in this study is secondary data obtained from Yahoo finance, dating from February 2006 to July 2022.

3.3 ARCH (q) Model

Understanding the history of ARCH model is necessary to transition to the GARCH (p,q) model. As presented in Sect. 1, [6], for his important contribution in modeling volatility was awarded the Noble Prize in 2003 for his contribution to modeling volatility and proposing the ARCH model. In the study, [6] illustrated that the series' mean, and variance can be simultaneously predicted. Conditional variance may be used as an AR process under ARCH. The model is given as (1).

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_n \varepsilon_{t-p}^2 \tag{1}$$

where $\varepsilon_t \sim iid(1, 0)$, so as to ensure the nonnegativity of the conditional variance, the parameters must satisfy the constraints and for where, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 0, 1, 2 \ldots p$ and the parameters are unknown.

3.4 GARCH (p, q) Model

An ARCH model extension was created by [4], enabling the conditional variance to be an ARMA. The resulting model is called a GARCH (p,q) model. For a log return series

(r_t), let the mean corrected return at time t be $\varepsilon_t = r_t - \mu_t$ [19]. The GARCH (p, q) model is represented by (2) and (3).

$$\varepsilon_t = \sigma_t z_t \tag{2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_1 \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_i \sigma_{t-j}^2 \tag{3}$$

where $z_t \sim (1, 0)$ and α_0 follows a Gaussian white noise. $\alpha_0 > 0, \alpha_1 \geq 0$ And $\sum_{j=1}^{\max(p,q)} \left(\alpha_i + \beta_j\right) \leq 1$ is required for the variance to be positive. Let $\alpha_i = 0$ for $i > p$ and $\beta_j = 0$ for $j > q$.

The GARCH (p, q) model can be easily simplified to an ARCH (p) model by equating p to zero. In this case conditional variance σ_t^2 changes over time, whereas unconditional variance exists when ε_t is present with $\sum_{j=1}^{\max(p,q)} \left(\alpha_i + \beta_j\right) < 1$.

Let $v_t = \varepsilon_t^2 - \sigma_t^2$, and when solving for σ_t^2 then the final equation becomes $\sigma_{t-i}^2 = \varepsilon_{t-i}^2 v_{t-i}$ for $i = (0, 1, 2 \ldots q)$. Using Eq. (3) and substituting σ_{t-i}^2 we can write GARCH (p,q) as (5).

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{\max(p,q)} \left(\alpha_i + \beta_j\right) \varepsilon_{t-i}^2 + v_{t-j}. \tag{5}$$

where $E(v_t) = 0$ and $\operatorname{cov}(v_t, v_{t-1}) = 0$ for $j \geq 1$. As a result, the series v_t is a martingale difference, rather than independent and identical distributed sequence (iid). In Eq. (2) it is shown that the GARCH model can be expressed as an ARMA process for the mean-corrected squared series [19].

3.5 ARIMA Process

The three main parts of an ARIMA model are $AR(p)$, which are the autoregressive terms or the order of lag observations in that model, then the difference between non-seasonal observations $I(d)$, and $MA(q)$, which is the size of the moving average window. The expression (p, d, q), is the order of an ARIMA model and the values represent the order or number of times the function occurs during the execution process of the model. In this case these values can be zeroes. The degree of $AR(p)$, difference (d), and $MA(q)$ in a $ARIMA(p, d, q)$ model can be stated mathematically as (5).

$$\begin{aligned} x_t = u_t + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} \\ - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned} \tag{5}$$

where:

(i) x_t is determined by differencing d times.
(ii) ε_t (hypothetical white noise) and in this case the assumption is that it is independently and identically distributed having a mean of zero and a variance which is constant σ_ε^2.

(iii) p and q are the number of autoregressive and moving average terms.
(iv) $\varphi_i (i = 1, 2, \ldots, p)$ and $\theta_i (i = 1, 2, \ldots, q)$ are the model parameters that are supposed to be estimated.

Parameter Estimation for the GARCH (p,q) Process

To estimate the parameter of the GARCH (p, q) process, the maximum likelihood method is used. The parameters to be estimated are $\alpha_0, \alpha_1, \ldots, \alpha_p$ and $\beta_1, \beta_2, \ldots, \beta_q$. . Moreover, the parameters are estimated in the very same way as in the ARMA (p, q) process. This study will maximize conditional likelihood, and we assume that both $\alpha_1, \alpha_2, \ldots, \alpha_p$ and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2$ are known. For this to occur, the assumption of normality must be satisfied. In that case the likelihood is given as (6).

$$\begin{aligned} & f\left(\varepsilon_1, \ldots, \varepsilon_t, \sigma_1^2, .., \sigma_t^2 | \theta\right) \\ & = f\left(\varepsilon_t, \sigma_1^2 | Y_{t-1}\right) f\left(\varepsilon_{t-1}, \sigma_{t-1}^2 | Y_{t-2}\right), \ldots, f\left(\varepsilon_2, \sigma_2^2 | Y_1\right) f(\varepsilon_1, \sigma_1^2 | \theta) \\ & = \prod_{t=2}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\varepsilon_t^2}{2\sigma_t^2}) f\left(\varepsilon_1, \sigma_1^2 | \theta\right). \end{aligned} \tag{6}$$

where θ and Y_{t-1} are the set of information at time $t - 1$ and f(ε$_1$, σ$_1^2$|θ) the probability density function is $f(\varepsilon_1, \sigma_1^2 | \theta)$. However, the exact form of $f\left(\varepsilon_1, \sigma_1^2 | \theta\right)$ is difficult to find, and we usually remove it from the prior likelihood function when the sample size is large enough.

3.6 Forecasting Using the GARCH (p,q) Process

When forecasting volatility of a return series using r_t the GARCH (p,q), the process is just like that of a GARCH (1,1) process. In this case we allow T to be the starting point for forecasting. A one-step ahead forecast for σ_{t+1}^2 is by (8).

$$\sigma_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{p} \left(\hat{\alpha}_i + \hat{\beta}_i\right) E[\varepsilon_{t+1-i}^2 | Y_t] - \sum_{i=1}^{q} \hat{\beta}_i E\left[V_{t+1-i} | Y_t\right] \tag{7}$$

where $\varepsilon_t^2 + \cdots + \varepsilon_{t+1-p}^2$ are known at time T. . Therefore, in a GARCH (p, q) model the k-step ahead forecast for conditional variance σ_{t+1}^2 is given as (8).

$$E\left[\varepsilon_{t+k}^2 | Y_t\right] = \hat{\alpha}_0 + \sum_{i=1}^{p} \left(\hat{\alpha}_i + \hat{\beta}_i\right) E[\varepsilon_{t+k-i}^2 | Y_t] - \sum_{i=1}^{q} \hat{\beta}_i E\left[V_{t+k-i} | Y_t\right] \tag{8}$$

where $E\left[\varepsilon_{t+k}^2 | Y_t\right]$ is recursive as given in the equation.

3.7 Model Selection

The selection criteria that were used and evaluated closely to check the accuracy of both models are the Maximum log Likelihood (ML) and the Akaike Information Criterion (AIC). The mentioned criteria are used to evaluate the forecasting efficiency and forecasting performance when comparing the models under study.

3.7.1 Akaike Information Criterion (AIC)

Akaike suggested the AIC criterion which uses the likelihood method. The approach comprises of estimating a range of potential ARMA models by the maximum likelihood method, and then calculating the AIC for each as presented in (9).

$$T.\log(SSR) + 2n \tag{9}$$

where the number of observations is given by T and the number of parameter estimates represented by n and this include the constant term. It can be noted that SSR can be reduced by Adding additional lags and the loss of degree of freedom that comes along with additional lags is penalized by these criteria.

3.7.2 Maximum Likelihood Estimation

To estimate the parameter, the maximum likelihood is employed in observing $Z_1, \ldots, Z_n$. From $ARIMA(p, d, q)$, the log likelihood on $Z_* = \left(Z_0, \ldots, Z_{1-p}\right)\alpha_* = \left(\alpha_0, \ldots, \alpha_{1-p}\right)$ is given as (10).

$$\begin{aligned} L\left(\phi, \theta, \sigma_\alpha^2\right) &= -\frac{n}{2} log\sigma_\alpha^2 - \frac{1}{2\sigma_\alpha^2}\sum_{t=1}^{n}\sigma_t^2 + C. \\ &= -\frac{n}{2} log\sigma_\alpha^2 - \frac{1}{2\sigma_\alpha^2} S(\phi, \theta) + C. \end{aligned} \tag{10}$$

where $\alpha_t = z_t - \phi_1 z_{t-1}, \ldots, \phi_p Z_{t-p} + \theta_1\alpha_{t-1} + \cdots + \theta_q\alpha_{t-q}$ from n large, one may choose to sum from α_{p+1}^2 onward and set previous $a's$ equal to 0. For q = 0, this yields the estimation of ϕ_j via the minimization of the least square scores.

$$\sum_{t=p+1}^{n}(Z_t - \phi_1 Z_{t-1}, \ldots, \phi_p Z_{t-p})^2.$$

For $ARIMA(p, d, q)$ $W_t = \nabla^d Z_t$.

4 Data Analysis

In this section, we present graphs, statistical tests, parameter estimation, model selection and the forecasting of the data done in the study.

4.1 Time Series Plot

The time series plot in Fig. 2 shows the pattern for standard bank closing prices from 2006 to 2022. The series follow a random walk process. The plot reveals that the series has a unit root due to the series' non-constant movement, implying that statistical properties like the mean and variance have been changing. Volatility clustering is apparent in Standard Bank, as shown by large and small price movements. The plot also highlights a significant peak in the stock prices during the year 2017. The plot also highlights a decline in 2020. This is due to the effects of Covid-19 on the South African economy.

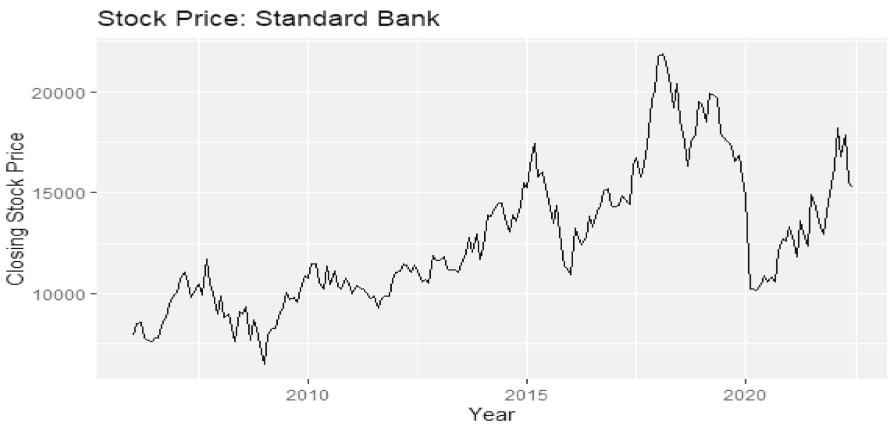


Fig. 2 Standard bank stock prices

4.1.1 Stationarity Testing for Original Time Series

The AUGMENTED DICKEY FULLER (ADF) test for stationarity for the the standard bank stock prices is presented in Fig. 3.

```
        Augmented Dickey-Fuller Test

data:  ts
Dickey-Fuller = -2.8334, Lag order = 5, p-value = 0.2272
alternative hypothesis: stationary
```

Fig. 3 ADF test for stationarity

Figure 3 shows that we cannot reject the null hypothesis because the *p*-value is not smaller than 0.05.This futher indicates that the time series is non-stationary. To put it another way, it has some time-dependent structure and does not exhibit constant variance over time. Therefore, differencing is shown below in Fig. 4.

The first differenced series in Fig. 4 above seems to fluctuate around the zero mean, implying that the series is stationary with respect to mean. This is further explained by the ADF below in Fig. 5.

The AUGMENTED DICKEY FULLER (ADF) test for stationarity for the first difference of the standard bank stock prices is presented below in Fig. 5.

The Augmented Dickey-Fuller (ADF) test in Fig. 5 investigate the unit-roots of the first difference series. This figure shows that the ADF is statistically significant at 5% level. This further indicates that the null hypothesis is rejected and conclude that the series is stationary, hence, it is mean reverting. The series is intergrated of order 1 or is I(1). The series has a unit root.

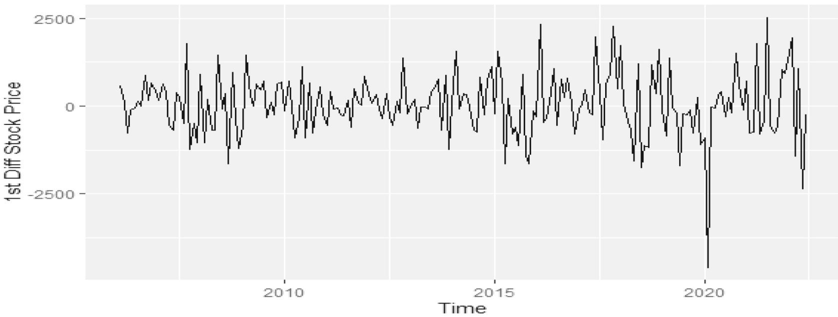


Fig. 4 First difference

```
        Augmented Dickey-Fuller Test

data:  diff((ts))
Dickey-Fuller = -14.658, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Fig. 5 ADF test for stationarity

4.1.2 Log Returns of Actual Series

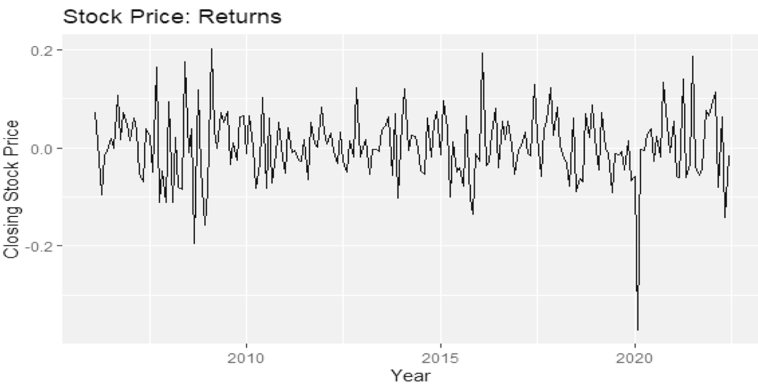


Fig. 6 Log returns

The daily returns time series plot demonstrates that statistical parameters such as variance or mean remain constant over time. The data is stationary as it revolves around the mean of 0. Figure 6 above depicts volatility clustering, which indicates a serial dependence in the time series data. Furthermore, a visual inspection of Standard Bank returns from 2006 to 2022, as shown in Fig. 6, reveals that volatility changes with time and that financial returns are clustered by these changes. This implies that small changes come after large changes, and large changes come after small changes. It can be observed that the returns fluctuate around the zero line, with the log return of stock prices occurring between 2016 and 2020. This signifies that the most volatile years were from 2016 to 2020.

4.1.3 Normality Test for Log Returns

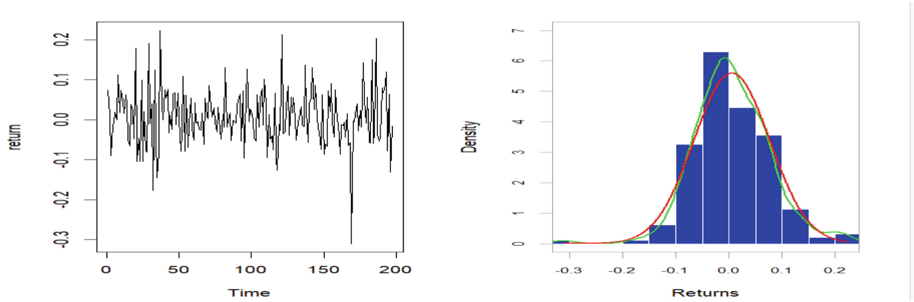


Fig. 7 Model rubustnes

Figure 7 shows a histogram of log returns which is normally distributed and symmetrical. Distribution of residuals are not different from a normal bell-shaped curve. The taller and slimmer body of the curve is a result of the log returns, as it exhibits excess kurtosis. The autocorrelation decrease rapidly over the lags, floating around 0.

4.2 Model Selection

Table 1 shows 10 permutations of model selection. Table 1 further demonstrate that GARCH order (1,1) and ARMA order (2,2) has the least AIC value. Therefore, it is the best model. With these model in hand now, we can use it to forecast for the next years.

4.3 Best Model GARCH(1,1)-ARMA(2,2)

The equation can be written as (Figs. 8 and 9):

$$\sigma_i^2 = 0.00 + 0.583x_{t-1} - 0.932x_{t-2} + 0.131\varepsilon_t - 0.637y_{t-1} + 1.013y_{t-2} + 0.735\sigma_{t-1}^2$$

4.4 Model Diagnostics

Figure 10 shows plots of ACF and PACF of the residuals respectively and based on this plots it shows that the residuals are stationary as the spikes of the ACF and PACF fall within the intervals. This is the indication that the chosen model GARCH(1,1)-ARMA(2,2) have managed to capture and model all the information in the time series. Furthermore Fig. 10 shows the normal q-q plot of the residuals which tests the normality assumption of the residuals and since the line of best fit satisfy the model, it can be concluded that the residuals are normally distributed.

Table 1 Model selection

Garch order	ARMA order	AIC
(1,1)	(0,0)	– 2.4104
(1,1)	(1,0)	– 2.4126
(1,1)	(1,1)	– 2.4139
(1,1)	(1,2)	– 2.3924
(1,1)	(2,0)	– 2.4139
(1,1)	**(2,2)**	**– 2.4507**
(1,2)	(1,1)	– 2.3924
(1,2)	(1,2)	– 2.3823
(1,2)	(2,0)	– 2.3926
(1,2)	(2,2)	– 2.3940

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(2,0,2)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.009867    0.005142    1.9189 0.055000
ar1     0.583271    0.013691   42.6011 0.000000
ar2    -0.932144    0.026808  -34.7709 0.000000
ma1    -0.637595    0.002793 -228.3163 0.000000
ma2     1.012975    0.004090  247.6693 0.000000
omega   0.000691    0.000515    1.3413 0.179813
alpha1  0.131133    0.108760    1.2057 0.227929
beta1   0.734638    0.172723    4.2533 0.000021

Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.009867    0.005035    1.9596 0.050038
ar1     0.583271    0.020649   28.2466 0.000000
ar2    -0.932144    0.029951  -31.1225 0.000000
ma1    -0.637595    0.004739 -134.5401 0.000000
ma2     1.012975    0.004189  241.8465 0.000000
omega   0.000691    0.000593    1.1641 0.244380
alpha1  0.131133    0.122613    1.0695 0.284850
beta1   0.734638    0.204385    3.5944 0.000325
```

Fig. 8 Parameter estimate for the best model

4.5 Forecast GARCH (1,1)-ARMA(2,2)

Figure 11 shows that the fitted model has a random behaviour.

Figure 12 shows that Standard Bank stock price will start to increase from the year 2028 with Highest Closing price of around 22100 and Lowest Closing price of around 15900.

```
LogLikelihood : 249.3941

Information Criteria
------------------------------------

Akaike       -2.4507
Bayes        -2.3174
Shibata      -2.4538
Hannan-Quinn -2.3967

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.5999  0.4386
Lag[2*(p+q)+(p+q)-1][11]   2.7276  1.0000
Lag[4*(p+q)+(p+q)-1][19]   7.8994  0.8080
d.o.f=4
H0 : No serial correlation
```

Fig. 9 Weighted Ljung-box test on standardized residuals

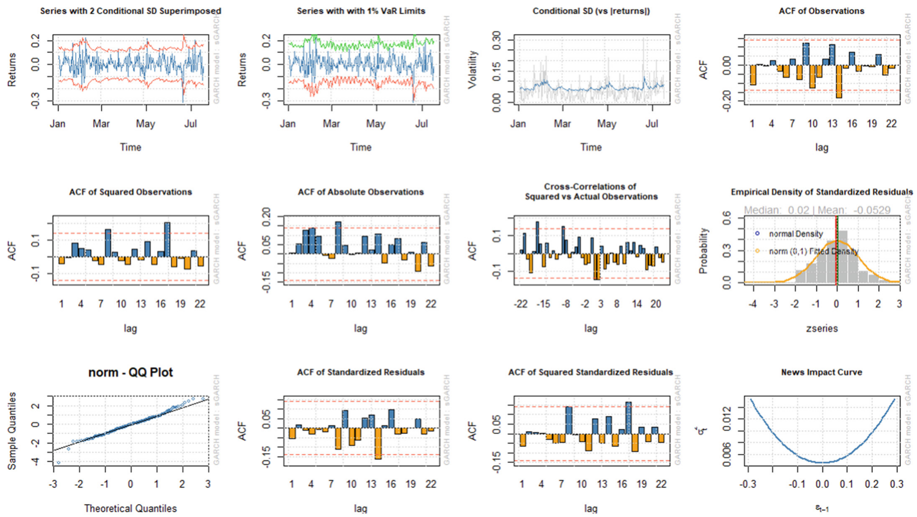


Fig. 10 Model diagnostics

5 Conclusions

In Sect. 4 the empirical results of 10 permutations of model selection were presented and the most efficient model was selected and then it was used for forecasting Standard Ban stock price. GARCH (1,1) and ARMA (2,2) were selected as the best models in modeling and forecasting volatility in the Standard Bank stock prices because it had the least AIC value of − 2.4507. The results showed that Standard Bank stock price will start to increase from the year 2023 with Highest Closing price of around 22100 and Lowest Closing price of around 15900. Exploring other models such as machine learning models in forecasting financial data is necessary.

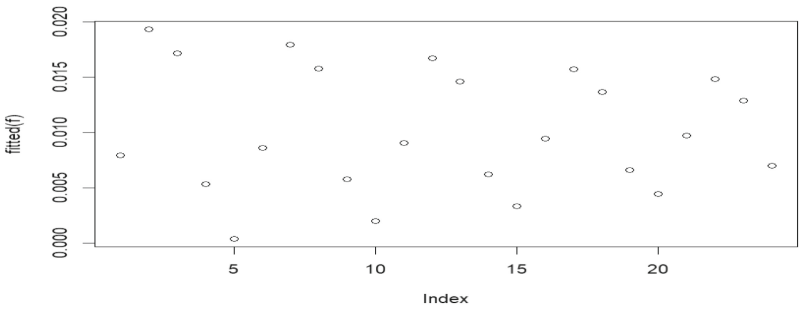


Fig. 11 Fitted model

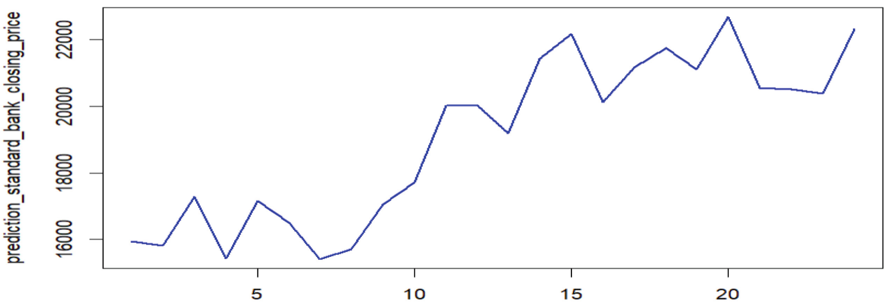


Fig. 12 Forecast for the next years

References

1. Alzyadat, J.A., Abuhommous, A.A., Alqaralleh, H.: Testing the conditional volatility of Saudi Arabia stock market: Symmetric and asymmetric autoregressive conditional heteroskedasticity (garch) approach. Acad Acc Finan Stud J **25**(2), 1–9 (2021)
2. Arnerić, J., Poklepović, T., Aljinović, Z.: Garch based artificial neural networks in forecasting conditional variance of stock returns. Croatian Oper Res Rev 329–343 (2014)
3. Bijelic, A., Ouijjane, T.: Predicting exchange rate value-at-risk and expected shortfall: A neural network approach (2019)
4. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**(3), 307–327 (1986)
5. Bouri, E., Jalkh, N., Roubaud, D.: Commodity volatility shocks and bric sovereign risk: A garch-quantile approach. Resour. Policy **61**, 385–392 (2019)
6. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica **50**(4), 987–1008 (1982). https://doi.org/10.2307/1912773.JSTOR1912773
7. Emenike, K.O.: Modelling stock returns volatility in nigeria using garch models (2010)
8. Fang, T., Lee, T.-H., Su, Z.: Predicting the long-term stock market volatility: A garch-midas model with variable selection. J. Empir. Financ. **58**, 36–49 (2020)
9. Hossain, A., Nasser, M.: Comparison of garch and neural network methods in financial time series prediction. In. 2008 11th International Conference on Computer and Information Technology. IEEE. pp. 729–734 (2008)

10. Hou, A., Suardi, S.: A nonparametric garch model of crude oil price return volatility. Energy Econ. **34**(2), 618–626 (2012)
11. Li, S., Luong, C., Angkola, F., Wu, Y.: Optimal asset portfolio with stochastic volatility under the mean-variance utility with state-dependent risk aversion. J. Ind. Manag. Optim. **12**(4), 1521 (2016)
12. Mala, R., Reddy, M.: Measuring stock market volatility in an emerging economy. Int. Res. J. Financ. Econ. **8**(5), 53–71 (2007)
13. Miah, M., Rahman, A.: Modelling volatility of daily stock returns: Is garch (1, 1) enough? Am. Acad. Sci. Res. J. Eng. Technol. Sci. **18**(1), 29–39 (2016)
14. Naik, M.S., Reddy, Y.: India vix and forecasting ability of symmetric and asymmetric Garch models (2021)
15. Raju, M., Ghosh, A.: Stock Market Volatility–An International Comparison. Securities and Exchange Board of India (2004)
16. Rastogi, S., Don, J., Nithya, V.: Volatility estimation using garch family of models: Comparison with option pricing. Pac. Bus. Rev. Int. **10**(8), 54–60 (2018)
17. Shaik, A., Syed, A.: Intraday return volatility in saudi stock market: An evidence from tadawul all share index. Manag. Sci. Lett. **9**(7), 1131–1140 (2019)
18. Taylor, J.W.: Generating volatility forecasts from value at risk estimates. Manage. Sci. **51**(5), 712–725 (2005)
19. Tsay, M., Brady, M.: A case study of cooperative learning and communication pedagogy: Does working in teams make a difference? J. Scholarsh. Teach. Learn. 78–89 (2010)
20. Van Den Hoogen, J., Meijer, S.: Lessons on the design of gaming simulations for convergence and divergence in volatile innovation environments. In: 2015 Winter Simulation Conference (WSC). IEEE. pp. 1044–1055 (2015)
21. Vasudevan, R., Vetrivel, S.: Forecasting stock market volatility using garch models: Evidence from the indian stock market. Asian J. Res. Soc. Sci. Humanit. **6**(8), 1565–1574 (2016)
22. Vejendla, A., Enke, D.: Evaluation of Garch, RNN and FNN models for forecasting volatility in the financial markets. IUP J. Financ. Risk Manag. 10(1) (2013)

Cebuano-English Code-Switching Speech Detection Using Support Vector Machine

Christian Dave O. Empinado(✉), Robine Cole A. Jumalon, and Christine F. Peña

University of San Carlos, P. del Rosario St, 6000 Cebu City, Philippines
{19100683,17100454,cfpena}@usc.edu.ph

Abstract. Code-Switching (CS) is a phenomenon where a person switches from one language to another in a single conversation. In the Philippines, CS is used often due to multiple languages that coexist together in a society as it is a simpler way of conveying messages between its members. This study will introduce a method that detects code-switching, specifically in the Cebuano-English language. A support vector machine (SVM) with language parameters from a Gaussian Mixture Model (GMM) was used for language classification. An Automatic Speech Recognition (ASR) system, consisting of Cebuano acoustic and language models, was developed to aid in the detection process. By combining language classification information and acoustic score, the method showed an increased accuracy of code-switching speech recognition compared to previous methods.

Keywords: Code-switching · Speech recognition · Language classification · Support vector machine

1 Introduction

Language is a conventional system that humans use to communicate with one another. Learning different languages can be done at school, at home, or in society. In communication, people often use specific languages that both the sender and the receiver can understand. However, when a person alternates between two different languages or more in a single conversation that is called code-switching [1]. Code-switching (CS) is a phenomenon that happens often to people who are exposed to two or more languages. Having multiple languages is difficult to manage, thus there are instances when one person may interchange the use of these languages during a conversation.

Over the years, Filipinos have evolved to use various Filipino languages as well as English together in sentences, this is how Filipinos developed the use of CS in conversation. For instance, in a group of peers, people often use CS as their means of communication because it is more comfortable for them to communicate. In some presentations, there are instances that the reporter might use CS to deliberate the topic well. English may not be the primary language for some Filipinos, so during instances when Filipinos converse with foreigners the act of using CS might be used involuntarily.

In some documentation and video blogs, there are moments where CS is used during recording. When this phenomenon occurs, the automatic caption of the video might

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 133–140, 2024.
https://doi.org/10.1007/978-3-031-50151-7_13

not always produce the expected output. These instances might cause some misunderstandings among viewers that are disabled. In some cases, friends often communicate using CS but not everyone is introduced to different languages, which may result in some confusion. Factors on language barriers and mistranslation may lead to miscommunication that may cause issues and misunderstandings. This study aimed to develop a Cebuano-English code-switching detection algorithm using SVM.

This study can help in gathering data for future linguistic research that involves code-switching in the region. Linguists may use the methods presented in this study in approaching code-switching studies and research. The study will provide advancements in the field of automatic speech recognition as well. This will introduce methods that are applicable to low-resource languages like Cebuano. The method can help in the improvement of captioning and transcription programs in the Cebuano language.

2 Related Work

2.1 Cebuano-English Code-Switching

Code-Switching (CS) has become a common phenomenon in the Philippines. It usually occurs when a person is exposed to different languages in school and society. Students undergo certain subjects that require them to use the English language and other subjects with the Filipino language. When engaging in a conversation, the languages that have been taught are used interchangeably. Every person has a reason they use CS. According to a survey, the majority of the students in the Senior High School of the University of Perpetual Help often use CS because it is comfortable, and they are able to express more on what they feel properly [2]. A similar study determined that Math teachers used Filipino-English CS most of the time to explain concepts in math in order for the students to understand the concepts they were teaching [3].

Cebuano-English Code-Switching (CECS) is a phenomenon where a person uses the Cebuano language and English language interchangeably in a single conversation. For instance, when a person says “Ganahan ko mu eat” there is a transition from Cebuano to English. Another example is “Study pa more” in literal English translation is “Study still more” this is a Filipino slang that cannot be translated into literal English, the original meaning of this word in English is “Keep Studying”.

CECS is often used in Cebu, Bohol, Davao, and other parts of Visayas. People from urban areas are more indulged in CECS than those from rural areas [4]. This means that people from urban areas are more likely to engage in conversations using CECS rather than strictly using only one language.

2.2 Automatic Speech Recognition (ASR)

Automatic Speech Recognition is a significant field in boosting human to computer interactions. There are a wide variety of studies done on ASR that complement this study. An acoustic and language model is needed for an ASR system to be viable and accurate. A study was done that uses Gaussian mixture modeling for detecting vowels in a given speech signal [5].

The Gaussian mixture model (GMM) is a very important statistical model for conventional speech recognition systems. This model may be used to represent a speech waveform into frame-based speech features [6]. However, another study states that Deep Neural Networks (DNN) performed much better in comparison to GMM on speech recognition tasks with large vocabulary volume [7].

The language model is responsible for predicting the next word in a sequence [8]. Most accurate language model is in the form of a Recurrent Neural Network (RNN). While RNN is one of the best forms in terms of accuracy, one study shows that this model may be improved to the point where an increase of more than 15 times has been observed in the speed of training and testing phases [9].

2.3 Problems and Approaches of Code-Switching Detection

Code-switching is prominent in many bilingual cultures, however, the majority of studies in the field of ASR focuses on monolingual speech recognition. Hence, data is scarce when it comes to code-switched speech since data on multiple languages are needed.

One of the major problems with code-switching detection is the usage of words that may or may not have a direct translation in the languages involved. A study proposed a language modelling approach in code-switching detection thus focusing on two tasks: the detection of code-switched sentences and the identification of code-switched words in a sentence. The language odelling approach has reached an accuracy rating of 79% [10].

Although data is readily available for most languages, data is scarce for Cebuano. This poses a problem as developing and training a code-switching detection algorithm requires the usage of enough mixed-language data in order to be accurate. However, a monolingual approach may be viable by training a code-switching language model using monolingual data. By using RNN models and constricting the output projection matrix, the study was able to bring embeddings of languages closer to each other. This approach is comparable to the approach of using synthesized code-switching data [11].

3 Experimental Setup

This study closely followed the method proposed by [12] where an SVM with GMM features is used for language detection together with ASR. This will be the basis of this study. With the help of an acoustic and language model, a speech audio is converted into a text transcript with language identification and code-switching times.

3.1 Corpus

Speech audio data is needed to train and test the necessary processes for the code-switching detection algorithm. The Cebuano and mixed corpus needed for the study was provided by the UP DSP laboratory which is licensed under the terms CC BY-NC-ND. This corpus was used in another study that features grapheme to phoneme transcription [13]. The English corpus that was used is taken from the CMU Arctic speech database.

3.2 Feature Extraction

Feature extraction is necessary for any automatic speech recognition process [14]. This process allows further processing of speech audio inputs in ASR systems. Mel Frequency Cepstral Coefficients (MFCCs) are the most common features in automatic speech recognition systems and may be used for code-switching detection [14].

The MFCC process begins with a speech audio divided into twenty millisecond audio frames. The "cepstral coefficients" are extracted from each frame [15]. The coefficients from each frame will form a feature parameter. This is used for Automatic Speech Recognition processes and code-switching detection. MFCCs feature extraction is done through the Librosa package in Python [16].

3.3 Acoustic Model

One of the key components of ASR is the ability to detect speech signals and process them. This is done by the acoustic model which takes a feature extracted speech audio, converts it to its respective phoneme equivalent and provides predictions at a given time [8]. The Cebuano acoustic model was trained using CMUSphinx to provide an output sequence of words. The English acoustic model from the pocketsphinx library was used.

3.4 Language Model

Another key component of ASR is the language model. The language model utilizes the output from the acoustic model. Using the N-gram rule, the model will predict the next syntactic word in a sequence of text. ASR systems may be feasible without a language model but are very limited in terms of accuracy [8]. A language model for Cebuano will be developed in order to provide an accurate text transcript of a given speech audio input. The English language model from the pocketsphinx library was used.

3.5 Gaussian Mixture Model

GMM models were used to aid in language classification. There are too many parameters if the extracted features are plugged into the SVM directly. Hence, a Gaussian Mixture Model (GMM) was used to establish a language feature. Instead of using the feature parameters in the SVM directly, it was first fitted in a GMM model and the corresponding log likelihood score of the features was taken per frame for input in the SVM. Two GMM models were developed. One for Cebuano using Cebuano speech data and one for English using English speech data. These were then labeled according to the language and then used as SVM training objects. A similar method was performed in another study [12].

3.6 Support Vector Machine

A Support Vector Machine is a supervised machine learning algorithm that is capable of classifying datasets. It is highly effective for smaller but complex datasets and requires little to no tweaking to improve accuracy. The aim of an SVM is to find a line or (n-1) dimension hyperplane that separates two classes [17].

In this study, the two classes are the languages English and Cebuano. In order to classify each data point into the two languages, a kernel function is needed. This study will utilize the Radial Basis kernel function also known as RBF. This kernel function is usually utilized with non-linear datasets [18]; hence it may be used for language classification that utilizes MFCC features [12]. The probability score was used to calculate the Language Likelihood Score for language classification. The scikit-learn library from python was used to establish the GMM and SVM models.

3.7 Language Classification

The steps in establishing the language classifier are as follows:

1. Feature extraction was performed, and a feature parameter was taken per sample.
2. The feature parameters were used to train a GMM model of the corresponding language. A GMM log-likelihood score was produced per sample.
3. The GMM log-likelihood scores were taken and used to train the SVM. The Cebuano GMM parameters were labeled as 1, while the English GMM parameters were labeled as 0.
4. For classification, a speech sample undergoes feature extraction. Two GMM language features are established. These features will then be classified by the SVM, and a probability score is produced for each feature.

3.8 Model Setup

The figure below depicts how the code-switching detection process goes. First, speech audio is taken as input. It will undergo feature extraction which outputs a feature parameter. The feature parameter will be decoded by the English and Cebuano decoder simultaneously (Fig. 1).

A GMM language feature is also established. The GMM language feature score undergoes classification through SVM with a probability score being produced. This is then combined with the acoustic model scores from each decoder. The higher combined score is taken as the result. Then the language model establishes the frame times. The following formula was used to determine the corresponding language score [12].

$$Score = 0.25 * Score_{SVM} + (1 - 0.25) * Score_{AM} \tag{1}$$

4 Results

4.1 Corpus Information

At The beginning of the experiment the approximate number of data collected from the corpus was 21 h. After further data cleaning the corpus was cut down to approximately 13 h, ¼ of the data was used for validation, and ¾ of the data was used for training. The training data was divided into 4 experiments in order to determine the best accuracy of the system (Table 1).

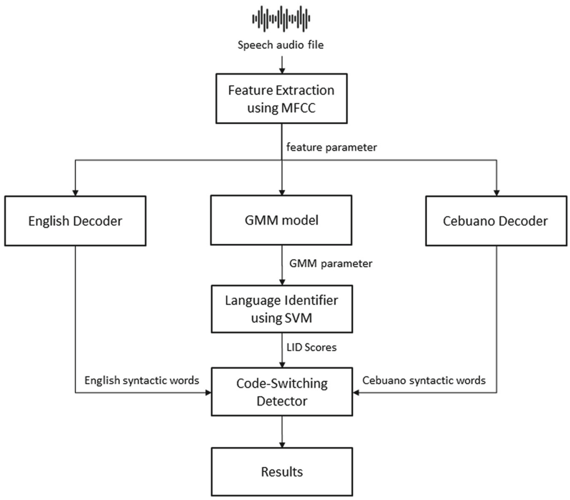


Fig. 1 Code-switching detection process.

Table 1 Corpus information for Cebuano.

Total corpus hours	Number of speakers	Number of audio files
10 h	71	15,256
7.5 h	52	11,086
5 h	34	7083
2.5 h	19	3848

4.2 Validation Method

Three metrics were chosen for validation: decoder accuracy, word error rate, and language classification accuracy. The accuracy is a comparison between the output transcript text with the correct given transcription text. The word error rate is the sum of deleted, inserted, and replaced words divided by the number of words. The language classification accuracy is also a comparison of the output language and the correct classified language.

4.3 Results

Findings from Table 2 showed that the longer the training hours the better the accuracy. However, due to inconsistencies in the data, the language classification accuracy reduced when total corpus hours were longer than 7.5 h. In this study, it was found that 7.5 h is the optimal training time to achieve the total best accuracy of the system while 2.5 h showed the minimum. Comparing this to the basis study, 58.4 h were used to train the system resulting in the highest language switching accuracy of 82.1%, an accuracy of 79.1%, and a word error rate of 22.7% [12].

Table 2 Cebuano decoder and code-switching detector results.

Total corpus hours	Cebuano decoder accuracy (%)	Word error rate (%)	Language classification accuracy (%)
10 h	81.8	10.1	66.2
7.5 h	**80.6**	**10.6**	**84.4**
5 h	79.6	10.8	71.9
2.5 h	79.8	11.5	55.7

5 Conclusion and Future Works

5.1 Conclusion

In this study, a method for Cebuano-English code-switching detection was developed through the use of an ASR system together with an SVM with GMM language feature inputs. Although data is not as abundant and available compared to previous studies, the proposed system was able to produce comparable results with previous studies through the validation methods seen in Sect. 115.2. The highest performing experiment was with 7.5 h of corpus data with a decoder accuracy of 80.6%, word error rate of 10.6%, and a language classification accuracy of 84.4%.

5.2 Future Works

For future works, the researchers would like to suggest increasing the number of corpus hours and expanding the vocabulary of the corpus to provide an opportunity for experimentation in improving the accuracy of the system. Furthermore, the researchers would like to suggest creating a code-switching corpus for a more proper validation of the system. A pure mono-language corpus is also needed for better model fitting. The methods presented in this study may also be used in other studies that involve other Filipino languages in code-switching speech like Tagalog, Ilocano, Hiligaynon, etc.

References

1. Salazar, D.: Switching Gears: Revising Code-Switching, n. Oxford English Dictionary. Retrieved 24 Apr 2022, from https://public.oed.com/blog/revising-code-switching/ (2020, Sept 21)
2. Rev.: All You Need to Know About Automatic Speech Recognition Transcription Models. Rev. Retrieved 13 Apr 2022, from https://www.rev.com/blog/guide-to-speech-recognition-transcription-models (2021, June 14)
3. Bravo-Sotelo, K.: Exploring the Tagalog-English code-switching types used for mathematics classroom instruction. IAFOR J. Educ. Lang. Learn. Educ. **8**(1). https://files.eric.ed.gov/fulltext/EJ1245827.pdf (2020)
4. Abastillas, G.: Divergence in Cebuano and English code-switching practices in Cebuano speech communities in the central Philippines A Thesis su. Retrieved 16 Apr 2022, from https://repository.library.georgetown.edu/bitstream/handle/10822/760907/Abastillas_georgetown_0076M_12963.pdf;sequence=1 (2015, Apr 1)

5. Kumar, A., Shahnawazuddin, S., Pradhan, G.: Exploring different acoustic modeling techniques for the detection of vowels in speech signal. In: 2016 Twenty Second National Conference on Communication (NCC), 1(2016), pp. 1–5 (2016). https://doi.org/10.1109/NCC.2016.7561195
6. Deng, L., Yu, D.: Automatic Speech Recognition: A Deep Learning Approach. Springer, London (2014)
7. Yu, D., Seltzer, M.L., Li, J., Huang, J.-T., Seide, F.: Feature learning in deep neural networks—studies on speech recognition tasks. Int Conf Learn Representations **1**(3), 9 (2013). arXiv:1301.3605
8. Republic of the Philippines: Mindanao Comprised About 24 Percent of the Philippines' Total Population. Philippine Statistics Authority. Retrieved 23 Apr 2022, from https://psa.gov.ph/content/mindanao-comprised-about-24-percent-philippines-total-population (2005, June 8)
9. Mikolov, T., Kombrink, S., Burget, L., Černocký, J., Khudanpur, S.: Extensions of recurrent neural network language model. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1(1), 5528–5531 (2011). https://doi.org/10.1109/ICASSP.2011.5947611
10. Yu, L.-C., He, W.-C., Chien, W.-N., Tseng, Y.-H.: Identification of code-switched sentences and words using language modeling approaches. Math. Probl. Eng. **2013**(1), 7 (2013). https://doi.org/10.1155/2013/898714
11. Chuang, S.-P., Sung, T.-W., Lee, H.-Y.: Training a code-switching language model with monolingual data. In: CASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020(1), 7949–7953 (2020, May). arXiv:1911.06003
12. Zhang, H.: Code-switching Speech Detection Method by Combination of Language and Acoustic Information. In: 2012 2nd International Conference on Computer and Information Application (ICCIA 2012), 1(1), pp. 372–375. https://www.atlantis-press.com/article/4038.pdf (2012)
13. Aquino, A., Tsang, J.L., Lucas, C.R., de Leon, F.: G2P and ASR techniques for low-resource phonetic transcription of Tagalog, Cebuano, and Hiligaynon. In: 2019 international symposium on multimedia and communication technology (ISMAC), 2019(1), pp. 1–5 (2019). https://doi.org/10.1109/ISMAC.2019.8836168
14. Afnan, S.: Comparison GMM and SVM classifier for automatic speaker verification. All Theses. https://tigerprints.clemson.edu/all_theses/2228 (2015)
15. Rao, K.S., Manjunath K.E.: Speech Recognition Using Articulatory and Excitation Source Features. Springer International Publishing (2017)
16. librosa.feature.mfcc — librosa 0.10.0.dev0 documentation. (n.d.). Librosa. Retrieved 25 July 2022, from https://librosa.org/doc/main/generated/librosa.feature.mfcc.html
17. Saini, A.: Support Vector Machine (SVM): A Complete guide for beginners. Analytics Vidhya. Retrieved 21 Apr 2022, from https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/ (2021, Oct 12)
18. Sreenivasa, S.: Radial Basis Function (RBF) Kernel: The Go-To Kernel. Towards Data Science. Retrieved 21 July 2022, from https://towardsdatascience.com/radial-basis-function-rbf-kernel-the-go-to-kernel-acf0d22c798a (2020, Oct 12)

Impact of Employability Skills on Employee Performance of Business Graduates in Nepal: Structural Equation Modeling Approach

Shrijan Gyanwali(✉), Bibek Karmacharya, and Ghanashyam Tiwari

School of Business, Pokhara University, 30, Pokhara, Nepal
gyanwalishrijan@gmail.com

Abstract. Employers have challenges merely to stay in business in today's highly competitive industry. To meet these challenges, they require workers with not only technical capabilities, but also employability qualities that allow them to adapt to a rapidly changing sector. The employability skills of business graduates are critical in assessing their potential to increase the organization's performance. The goal of this study is to analyze and factorize the components and dimensions of employability abilities that impact the employability and performance of business graduates. The study adopts a descriptive survey-based research approach, in which data are collected from a sample of respondents using a self-administered survey questionnaire. The convenience sampling method is used to obtain a sample size of 422 respondents. The statistical techniques used for data analysis are Descriptive Statistics, Confirmatory Factor analysis and Structural Equation Modeling. The results of the Confirmatory Factor Analysis and Path Analysis showed that critical thinking and problem-solving skills, technology skills, organizational skills, and continuous learning skills had a significant impact on employee performance. The study highlights the importance of these skills in enhancing employee performance and the need for organizations to focus on developing these skills in their employees. The findings of this study can be useful for human resource managers and organizations in developing training programs to improve the skills and performance of their employees.

Keywords: Employability skills · Employee performance · Structural equation modeling

1 Introduction

Employers face obstacles in today's highly competitive market just to stay in business. To face these difficulties, they need individuals who not only have technical abilities, but also employability skills that allow them to adapt to a quickly changing industry [1]. Employability skills are the characteristics, abilities, and information that enable people to be work-ready and effective in the job [2]. They assist individuals in adapting to changes in the workplace and advancing their careers. Business graduates play an important part in an organization's success [3]. Their employability abilities are crucial in determining

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 141–152, 2024.
https://doi.org/10.1007/978-3-031-50151-7_14

their capacity to improve the organization's success [4]. Companies are searching for graduates with broad abilities, such as good communication, problem-solving, interaction skills, self-initiative, and efficiency, in addition to discipline-specific talents [5]. Mindset and personal attributes such as loyalty, dedication, honesty, consistency, and integrity are also important in determining graduates' employment [2].

In spite of the focus on employability skills, there remains a disparity between what is learned throughout the study and what is expected by companies [6]. The issue of joblessness in Nepal and throughout the world is caused mostly by a shortage of qualified graduates. Academic institutions are responsible for providing graduates with employability skills, yet the skills, practices, and mindsets required by employers differ from what is taught during the study [7]. As a result, it is necessary to define the precise employability skills that employers expect from business graduates [3]. The purpose of this research is to investigate and factorize the components and dimensions of employability abilities that influence the employability of business graduates and their performance. The research helps academic institutions in capturing difference in employability skills as they relate to employment requirements and in producing excellent business graduates.

2 Literature Review

Good communication skills, such as verbal and nonverbal communication, listening, and understanding people, are essential for establishing connections and attaining goals in the workplace [3]. Employers place a high emphasis on personal abilities like communication, problem-solving, and critical thinking since they contribute favorably to office culture and the company's performance [5]. Personal skills, unlike practical talents, are elusive abilities that help people to successfully express ideas, project a pleasant attitude at work, and fulfill commitments, leading in a focused and efficient workforce [8].

In company, teamwork entails a group of individuals who collaborate to accomplish a collective objective [5, 9]. When a group works successfully together, they can achieve more than individuals working alone since varied abilities can result to more sensible solutions. Successful leaders must enable communication and cooperation, effectively allocate work, and give direction and support when necessary [8]. Problem-solving and critical thinking are crucial abilities that refer to the capacity to solve issues using information, facts, and data. Employees has to be able to assess events, recognize issues, and find effective solutions in the workplace, therefore both talents are essential [10].

Businesses look for employees that can help them harness current technologies to stay ahead of the competition [11]. Depending on the work, technological abilities may range from simple word processing and email to specialized program design tools and visual effects [12]. Learning and practice are two ways to improve your IT skills. Good organizational abilities are also required, which include the ability to identify tasks, prioritize them, create timetables, and complete them on time [7]. Constant learning is essential for staying relevant in the workplace, and businesses may encourage a learning culture by providing training programs, online resources, and mentorship opportunities [13].

Employee performance is an indicator of an employee's capacity for carrying out work obligations in an effective, efficient, and high-quality manner [14]. Good communication skills, excellent problem-solving and critical thinking abilities, effective time management, flexibility, and leadership abilities may all have a significant impact on an employee's effectiveness [15]. Companies should focus staff skill development and enhancement in order to increase overall performance and aid in the achievement of their organization. Business graduates can boost their prospects of success in the service industry by concentrating on organizational skills, continuous learning, and employee performance [10].

Theoretical Framework

Based on previous studies and conceptual review the theoretical framework of the study is as follows (Fig. 1).

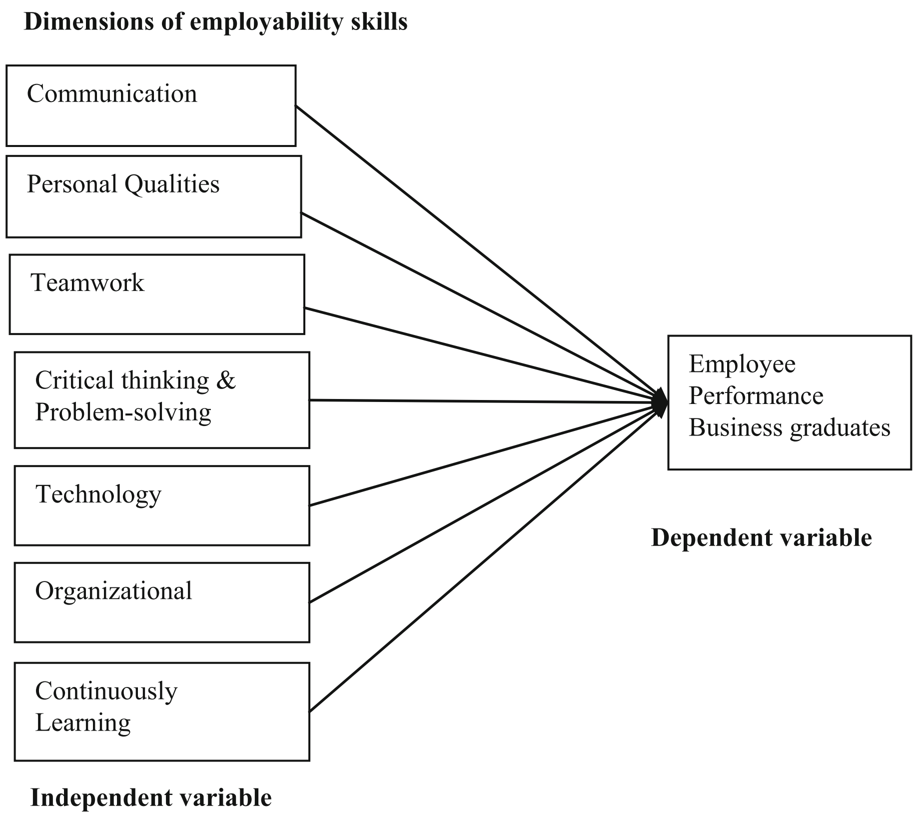


Fig. 1 Theoretical framework of showing the impact of various employability skills on employee performance. *Source* [16, 17]

Research Gap: While there have been studies on employability skills and their impact on job performance, there is a lack of research specifically focused on business graduates in Nepal. This knowledge gap highlights the need for further research on the specific employability skills required by Nepalese business graduates and their impact on employability. Therefore, the proposed study aims to address this gap by exploring the constructs and dimensions of employability skills that affect the employability of business graduates in Nepal. The study will contribute to the existing literature on employability skills by filling the gap in knowledge related to the Nepalese context and providing insights for academic institutions and employers on how to improve the employability of business graduates.

3 Instruments and Methods

This research study is a quantitative research study aimed at analyzing the impact of employability skills of business graduates on their employability. The study employs a descriptive survey-based research design, which involves collecting data from a sample of respondents using a self-administered survey questionnaire. The population for this study is all business employers working in different business sectors and firms within Pokhara valley. Convenience sampling technique is used to acquire the sample of 422 which is greater than desired sample of 384 at 5 percent level of significance and error margin [18].

The survey questionnaire consists of 42 observed variables, which measure the impact of employability skills of business graduates on their employability. The questionnaire is designed based on the theoretical models from previous studies, and indicators for measurement are supposed to reflect truly the concept [19] of employability skills which is represented in Appendix. In this study, the validity of the measures is assessed using convergent validity and discriminant validity. Convergent validity is assessed using Average Variance Extracted (AVE), which measures the degree to which the variance of the indicators is accounted for by the construct [20]. Discriminant validity is assessed using the Fornell-Larcker criterion, which compares the AVE of each construct to the squared correlations between the constructs [20]. In this study, the reliability of the measures is assessed using Cronbach's alpha and composite reliability [21].

After collecting the data through a survey, it is entered and analyzed in Statistical Package for Social Science (SPSS version 26) and AMOS (version 22). The statistical techniques used for data analysis: Descriptive Statistics, Cronbach's Alpha Test, Confirmatory Factor analysis and Structural Equation Modeling. Participants were informed about the purpose of the study, the procedures involved, and their rights as participants. Informed consent was obtained from all participants, and their confidentiality and anonymity were ensured. Participants were informed that they could withdraw from the study at any time without any negative consequences [22].

4 Analysis of Results and Discussion

4.1 Demographic Analysis

Descriptive analysis consists of 442 respondents who belong to different gender, age, work experience, monthly income, and nature of service. Among the total respondents, 283 (64%) are male, and 159 (36%) are female. This implies that there is a higher representation of male employers in the service industry of Pokhara valley. The age of the respondents ranges from below 20 to above 50 years. The largest group of respondents (40.7%) falls between 20–30 years, followed by 33% between 30–40 years. The smallest group of respondents is above 50 years (4.5%). This indicates that most of the employers in the service industry of Pokhara valley belong to the younger age group. Among the respondents, 267 (60.4%) are married, 169 (38.2%) are unmarried, and only 6 (1.4%) have specified their marital status as 'others.' This indicates that the majority of the employers in the service industry of Pokhara valley are married. Regarding work experience, nearly a quarter of the respondents (23.5%) have less than three years of experience, while 26.9% have over ten years of experience. 99 (22.4%) respondents have 3–5 years of experience, and 120 (27.1%) have 5–10 years of experience. This suggests that the employers in the service industry of Pokhara valley have diverse work experience. The majority of the respondents (35.1%) have a monthly income between Rs 20,000 and 40,000. 112 (25.3%) respondents have an income above Rs 60,000, while 111 (25.1%) have an income below Rs 20,000. This implies that the income of the employers in the service industry of Pokhara valley is diverse. Regarding their association with organizations, the largest groups of respondents (32.4%) have been associated with their organization for 2–5 years, followed by 25.1% having less than two years of association. Only 13.8% have been associated with their organization for above 12 years. This suggests that the employers in the service industry of Pokhara valley have diverse association with their organizations. The respondents belong to different service sectors such as banking and financial institution (37.3%), tourism and hospitality (31.2%), education and consultancy (19.9%), and health services (11.5%).

4.2 Reliability and Validity

The reliability and validity of a research approach are critical to ensuring the trustworthiness and accuracy of its findings. In this study, construct reliability was evaluated using Cronbach's Alpha (α) and Composite Reliability (CR), with a reliability threshold of 0.70 [21]. The results from Table 1 indicate that all Cronbach's Alpha values exceeded this threshold, ranging from 0.715 for Communication skills to 0.839 for Continuous learning skills. Similarly, the model's Composite Reliability exceeded the threshold, with Communication skills having the lowest value at 0.744 and Technology skills having the highest value at 0.866. Overall, these results suggest that the construct is reliable.

In addition to construct reliability, construct validity was assessed using the subcategories of discriminant and convergent validity. The results from Tables 1 and 2 indicate that the model met the proposed criteria of CR > 0.70 and AVE > 0.50 [21]. Furthermore, the proposed requirements of Mean shared value (MSV) smaller than Average variance explained (AVE) and Square Root of AVE greater than Inter-construct correlation were fulfilled, confirming discriminant validity [20].

4.3 Confirmatory Factor Analysis (CFA)

The results of the confirmatory factor analysis (CFA) show that the measurement model has a good fit based on various fit indices. The study used structural equation modeling in AMOS and maximum likelihood estimate [23]. The fit indices include GFI and AGFI, which indicate a good fit with values of 0.932 and 0.929, respectively. However, sample

Table 1. Construct reliability of scale

Variables	α	CR	AVE	MSE
Communication skill	0.715	0.744	0.606	0.523
Personal qualities skill	0.753	0.785	0.635	0.594
Teamwork skill	0.782	0.828	0.697	0.632
Critical Thinking skill	0.824	0.861	0.704	0.661
Technology skill	0.807	0.866	0.697	0.645
Organizational skill	0.812	0.854	0.752	0.701
Continuous learning skill	0.839	0.852	0.735	0.715

Source Field Survey, 2022 and authors' calculation

Table 2. Square root of AVE and construct correlation analysis

Variables	Com	Per	Tea	Cri	Tec	Org	Con
Com	**0.778**						
Per	0.318	**0.797**					
Tea	0.522	0.365	**0.835**				
Cri	0.459	0.542	0.427	**0.839**			
Tec	0.382	0.593	0.349	0.389	**0.835**		
Org	0.438	0.513	0.402	0.286	0.457	**0.867**	
Con	0.516	0.509	0.491	0.436	0.398	0.483	**0.857**

Source Field Survey, 2022 and authors' calculation

size may affect these indices. The average chi-square ($\chi 2$) value is 1.970, which falls short of the 3.0 threshold but can be biased [24]. The RMSEA is 0.037, which is less than the 0.08 threshold and indicates a good fit. The CFI and TLI are 0.953 and 0.959 respectively, which are considered incremental fit indices and show a good fit. In summary, the CFA results indicate that the measurement model has a good fit, with several fit indices supporting this conclusion (Fig. 2).

4.4 Structural Model or Path Analysis

The emphasis shifts from the relationships between latent constructs and observable variables to the nature and magnitude of the linkages between the constructs when the research proceeds from the measurement model to the structural model, as seen in the Fig. 3. Utilizing well-known economic theories, the structural model is defined. It is postulated that the Communication skills, Personal qualities skills, Teamwork skills, Critical thinking skills, Technology skills, Organizational skills and Continuous learning skills propositioned to Employee performance of business graduates of various service sectors within Pokhara. The SEM path analysis findings are provided in the Table 3. The

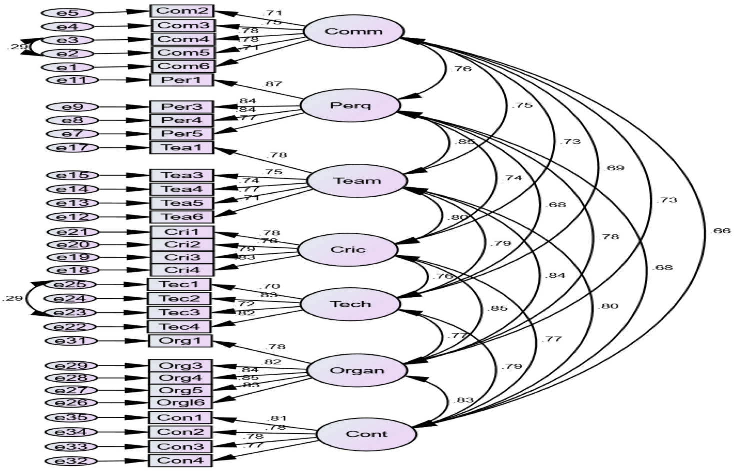


Fig. 2. Measurement model of the employability skills of business graduates

research investigates the postulated causal link presented in the theoretical model using structural or path analysis.

Table 3. SEM path analysis

Structural path	Estimate	SRW[a]
Employee performance ← Communication	0.157 (0.103)	− 0.126
Employee performance ← Personal qualities	0.118 (0.124)	0.098
Employee performance ← Teamwork	− 0.035 (0.181)	− 0.027
Employee performance ← Critical thinking	0.311** (0.124)	0.268
Employee performance ← Technology	− 0.217** (0.109)	− 0.189
Employee performance ← Orgnizational	0.476*** (0.137)	0.430
Employee performance ← Continuous learning	0.279** (0.131)	0.223

Source Field Survey, 2022 and authors' calculation
Note Value in parentheses () indicate *p*-value

The path analysis findings display the overall fit measures stated in the previous section, which offer judgment on how well the structural or path model matches the data. The analysis of path model outputs reveals that $\chi 2$ (635) = 1235.486 with *p* value

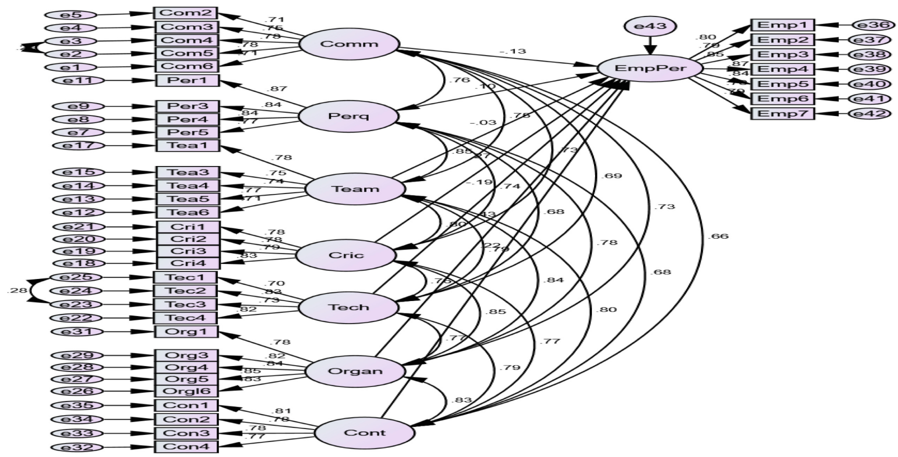


Fig. 3. SEM path analysis indicating relationship between employability skills and employee performance

= 0.000, GFI = 0.913, AGFI = 0.901, CFI = 0.951, TLI = 0.946, RMSEA = 0.039 produced a excellent model fit. The results show that communication skills, personal qualities skills, and teamwork skills have negative and insignificant impacts on employee performance, indicating a lack of alignment between these skills and job requirements or poor communication and team dynamics within the organization. In contrast, critical thinking, organizational skills, and continuous learning skills have positive and significant impacts on employee performance. The relationship between organizational skills and employee performance is the highest and most significant. The values of squared multiple correlations reveal that around 67% of variation in Employee performance is explained by the combine effect of independent variable. Overall, the analysis suggests that business graduates may possess the right skills, but they need to align with the job requirements to translate into improved performance, and employees need to continually learn and adapt to stay competitive in the service sector.

4.5 Discussion

The study presents an overview of the respondents' demographic characteristics, work experience, association with the organization, and the nature of service, along with the scores obtained for various skills and their association with different variables. The majority of respondents was male and had higher educational backgrounds. The study used Confirmatory Factor Analysis and Path Analysis to assess the fit of the measurement and structural models, respectively, and found that both models had an excellent fit. Hypothesis testing showed that critical thinking skill, technology skill, organizational skill, and continuous learning skill had a significant impact on employee performance, while communication skill, personal qualities skill, and teamwork skill had an insignificant impact. The R square value indicated that the combination of all employability skills explains a significant portion of the variation in employee performance. The study findings can help organizations to identify the strengths and weaknesses of their employees and design effective training programs. Organizations can also use the findings to recruit employees based on their skill sets and match their skills with the nature of service. Finally, the study also provides insights for future research on the relationship between demographic variables, skills, and organizational outcomes.

5 Conclusion

Based on the findings of the study, it can be concluded that the majority of the respondents were males with a higher educational background and diversified work experience. The respondents were associated with different types of organizations, with the highest number of respondents associated with banking and financial institutions. The descriptive statistics for different skills revealed that Effective Speaking Skills and Responsibility received the highest mean scores, while Effective Writing Strategies and Self-Esteem received the lowest mean scores. The results of the confirmatory factor analysis and path analysis showed that majority of the independent variable like critical thinking and problem-solving skills, technology skills, organizational skills, and continuous learning skills had a significant impact on employee performance. The study highlights the importance of these skills in enhancing employee performance and the need for organizations to focus on developing these skills in their employees. The findings of this study can be useful for human resource managers and organizations in developing training programs to improve the skills and performance of their employees.

Implication and Direction for Future Research

The study's findings shed light on the characteristics and skills of employees across various sectors, but more research is needed to understand the factors that influence the development of these skills, such as training and work environment. Future research should focus on identifying effective training programs and assessing their impact on employee performance, as well as exploring the relationship between employee skills, job satisfaction, and turnover intention. It would also be beneficial to examine how these skills vary in different cultural and organizational contexts. Ultimately, continuing research on employee skill development and utilization in different sectors can help improve organizational effectiveness and productivity.

Acknowledgements. This research is funded by University Grants Commission, Nepal.

Appendix

Operationalization of Study Variables

A) Dimension of Employability Skills

Communication Skill
1. Effective Reading Strategies
2. Effective Writing Strategies
3. Using numeracy effectively
4. Effective Listening Skills
5. Effective Speaking Skills
6. Share information using ICT

Team Work Skill
1. Work independently or as part of a team
2. Coaching and mentoring skills
3. Serves Clients or Customers
4. Exhibits Leadership
5. Flexibility
6. Works with Cultural Diversity

Personal Qualities Skill
1. Responsibility
2. Self- Esteem
3. Sociability
4. Honesty and high integrity
5. Self- Management

Critical Thinking & Problem Solving
1. Applies creative & innovative solutions
2. Applying practical approach for solution
3. Applies decision-making strategies
4. Recognizes and Solves Problems

Technology Skill
1. Selects Technology
2. Applies Technology to Task
3. Maintain & troubleshoot technology
4. Having information technology skill

Continuous Learning Skill
1. Having enthusiasm for ongoing learning
2. Managing own learning
3. Accessibility of the learning resources
4. Assess personal strengths and areas for development

Organizational Skill
1. Manages Time
2. Manages Money
3. Manages Materials/Facilities
4. Planning Process
5. Adapt to changing requirements & information
6. Continuously monitor the success of a project or task and identify ways to improve

B) Employee Performance
1 Leading the team in achieving the goals
2 Developing subordinates and other
3 Customer satisfaction oriented
4 Agility/Adaptability in the organization
5 Judgement capabilities
6 Attitude and emotional stability
7 Initativeness

References

1. Asefer, A., Abidin, Z.: Soft skills and graduates' employability in the 21st century from employers' perspectives: a review of literature. Int. J. Infrastruct. Res. Manag. [Internet].**9**(2):44–59. Available from: https://iukl.edu.my/rmc/publications/ijirm/ (2021)

2. Mainga, W., Braynen, M.B.M., Moxey, R., Quddus, S.A.: Graduate employability of business students. Adm. Sci. **12**(3) (2022)
3. Milon, M., Islam, M.R., Khan, A.R.: Employability skills of business graduates in job markets of Bangladesh: investigating weaknesses and suggesting improvements. Int J Manag Account. **3**(1), 20–27 (2021)
4. Andrews, J., Higson, H.: Graduate employability, "soft skills" versus "hard" business knowledge: a European study. High Educ Eur. **33**(4), 411–422 (2008)
5. Pinto, L.H., Ramalheira, D.C.: Perceived employability of business graduates: the effect of academic performance and extracurricular activities. J Vocat Behav [Internet] **99**, 165–78 (2017). https://doi.org/10.1016/j.jvb.2017.01.005
6. Buntat, Y., Jabor, M.K., Saud, M.S., Mansor, S.M.S.S., Mustaffa, N.H.: Employability skills element's: difference perspective between teaching staff and employers industrial in Malaysia. Procedia—Soc Behav Sci [Internet]. **93**(1990), 1531–5 (2013). https://doi.org/10.1016/j.sbspro.2013.10.077
7. Gaite, F.T.G., Arroyo, R.A., Lim, P.J., Vergara, P.M.E., Doria, J.A.: Employers' feedback on job performance of hotel and restaurant management graduates. Int. J. Acad. Ind. Res. **3**(4), 129–144 (2022)
8. Khattak, A.N., Irfan. K.U., Karim, A.: Employability Skills: Career Challenges for Fresh Bachelor of Business Administration (BBA) Graduates in Pakistan, vol. 7, issue 1, pp. 490–503 (2023)
9. Amen, U.: Employer's expectations versus performance of fresh graduates: Business schools. Mark Forces Coll. Manag. Sci. **IX**(2), 39–50 (2014)
10. Mainga, W., Daniel, R.M., Alamil, L.: Perceptions of employability skills of undergraduate business students in a developing country: an exploratory study. High. Learn. Res. Commun. **12**(1), 28–63 (2022)
11. Hodges, D., Burchell, N.: Business graduate competencies: employers' views on importance and performance. J. Coop. Educ. **2003**(4), 16–22 (1993)
12. Mehrotra, A., Elias, H.: Employers' feedback on business graduates. US-China Educ. Rev A. **7**(4), 190–199 (2017)
13. Fajaryati, N., Budiyono, B., Akhyar, M., Wiranto, W.: The employability skills needed to face the demands of work in the future: Systematic literature reviews. Open Eng. **10**(1), 595–603 (2020)
14. Tanius, E., Johari, H. bt, Yulia, A., Siong, H.C., Pazim, K.H.: The employability skills performance of business graduates in Malaysia: do employers, graduates and academicians speak the same language? Int. J. Asian Soc. Sci. **9**(1), 11–7 (2019)
15. Hosain, M.S., Mustafi, M.A.A., Parvin, T.: Factors affecting the employability of private university graduates: an exploratory study on Bangladeshi employers. PSU Res Rev. (2021)
16. Abas, M.C., Imam, O.A.: Graduates' competence on employability skills and job performance. Int J Eval Res Educ. **5**(2), 119 (2016)
17. Rasul, M.S., Abd Rauf, R.A., Mansor, A.N., Puvanasvaran, A.P.: Employability skills assessment tool development. Int Educ Stud. **5**(5), 43–56 (2012)
18. Singh, A.S., Masuku, M.B.: Department. Sampling techniques and determination of sample size in applied statistics research. Inwood Mag. **II**(96), 32–33 (2011)
19. Shepard, L.A.: Chapter 9: Evaluating Test Validity. Rev. Res. Educ. **19**(1), 405–450 (1993)
20. Farrell, A.M.: Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). J. Bus. Res. [Internet]. **63**(3), 324–7 (2010). Available from: https://doi.org/10.1016/j.jbusres.2009.05.003
21. Şimşek, G.G., Noyan, F.: McDonald's ω t, Cronbach's α, and generalized θ for composite reliability of common factors structures. Commun. Stat. Simul. Comput. **42**(9), 2008–2025 (2013)

22. Miller, T., Boulton, M.: Changing constructions of informed consent: Qualitative research and complex social worlds. Soc. Sci. Med. **65**(11), 2199–2211 (2007)
23. Mîndrilă, D.: Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. Int. J. Digit. Soc. **1**(1), 60–66 (2010)
24. Jackson, D.L., Gillaspy, J.A., Purc-Stephenson, R.: Reporting practices in confirmatory factor analysis: An overview and some recommendations. Psychol. Methods **14**(1), 6–23 (2009)

Modelling and Forecasting South African Airline Passengers Using ARIMA

Thato Theo Maruping[1], Phemelo Seaketso[1], Precious Mdlongwa[2], and Elias Munapo[1](✉)

[1] Department of Business Statistics and Operations Research, School of Economic Sciences, North West University, Mafikeng Campus, Mafikeng, South Africa
elias.munapo@nwu.ac.za

[2] National University of Science and Technology, PO Box AC 939, Ascot, Bulawayo, Zimbabwe

Abstract. In the study we model and forecast South African airline passengers, using yearly time series data from 1970 to 2019. Box-Jenkins ARIMA methodology is used to forecast airline passenger data for the next five years (2020–2024). The raw data used was not stationary and as a result, was differenced once to make it stationary. In this study, an ARIMA (3,1,2) model was proposed. In addition, diagnostic tests demonstrated that the proposed model is, in fact, adequate and could be utilised to make predictions on the number of passengers for the South African airline from 2020 to 2024.

Keywords: Forecasting · Modelling · SA airline passengers · ARIMA · Box · Jenkins

Abbreviations

The following abbreviations and what they stand for are used.

AFC	Autocorrelation Function
AIC	Akaike's Information Criterion
AR	Autoregression
ARIMA	Autoregressive Integrated Moving Average
ESACF	Extended Sample Autocorrelation Function
GDP	Gross Domestic Product
I	Integrated
MA	Average
PACF	Partial Autocorrelation Function
Q-Q	Quantile-Quantile plot
SAA	South African Airline
SBC	Schwarz's Bayesian Criterion
TBATS	Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trends, Seasonal components
WDI	World Development Indicators

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 153–161, 2024.
https://doi.org/10.1007/978-3-031-50151-7_15

An impressive annual average rate of 6.31 percent for airline passengers was recorded in South Africa. According to data from the World Development Indicators, South Africa saw an increase from roughly 1.52 million travellers in 1970 to 25.7 million travellers in 2019. (WDI). Although this demonstrates a significant improvement, South African Airlines (SAA) is no different from other businesses in that it requires proper planning for the short and long term. Understanding the anticipated demand is necessary for effective planning so that each client's needs can be met. To put it another way, it is necessary to comprehend the past in order to forecast the future. It is crucial to accurately predict this demand because there is a greater demand for airline passengers in South Africa.

A method of statistical analysis known as the Autoregressive Integrated Moving Average (ARIMA) makes use of time series to learn about and comprehend data in order to forecast or predict the future. The ARIMA model is autoregressive in the sense that it predicts the future based on past data. In this paper, we develop an ARIMA model to forecast future SA airline passenger demand in South Africa from 2020 to 2024. Box and Jenkins (1970) provided us with procedures for developing forecasting models.

1 Some Similar Studies Done

Nilgun and Betul (2016) investigated the use of data from neighbouring airports in forecasting an airport's traffic volume, as well as the impact of neighbouring countries on air passengers from a domain country. They also contrasted trigonometric seasonality, Box-Cox transformation ARMA errors, trend, seasonal components (TBATS), and ARIMA models to reflect the performance of neighbouring dependent countries. They discovered that TBATS models outperform ARIMA models and plan to use them for seasonal long-term airline passenger data.

The ARIMA (2,1,0) model was found to be the most effective in research to forecast sugarcane production in India over the next five years. According to Minoj and Madhu (2014), when the model was statistically evaluated, the model residuals were found to be normally distributed with a mean of zero and a constant variance.

Jamal et al. (2018) utilised modelling and forecasting techniques to predict future demand using a time-series approach. The ARIMA (1,0,1) model was found to be the most accurate in this investigation that was carried out for the benefit of food companies. The model can be used by those in the food manufacturing industry to forecast demand and make informed business decisions.

Abonazel and Abd-Elftah (2019) developed an ARIMA model for Egypt's gross domestic product (GDP). The GDP was modelled and forecasted using data from 1965 to 2016. The World Bank dataset and the ARIMA model (1,2,1) were statistically sufficient for predicting Egypt's GDP for the following ten years.

2 Research Methodology

2.1 Data

The data set utilised for this analysis spans the years 1970 to 2019. The source of annual data on South African airline passengers was obtained from.

2.2 ARIMA Model

The ARIMA model is composed of three parts. The first component is the autoregression (AR) model, which deals with the regression of a changing variable on its lags. The second component is the integrated (I), which is the stage at which the non-stationarity of the data is transformed into stationarity. The moving average (MA), which incorporates the residual error and the moving average to test for dependence on lag-timed observations, is the third and final component. The acronym for it is generally ARIMA (p, d, q), where p stands for AR, d for integration, and q for MA (Ming et al. 2014). An ARIMA model thus shows that the data can be non-stationary but can be differentiated to become a stationary process, enabling forecasting. Therefore, it might be useful for modelling and predicting South African airline passengers over the following five years (Jansson and Larsson 2019).

ARIMA (p,d,q) model is given by Eq. (1)

$$\left(1 - \varphi_1 B - \ldots - \varphi_p B^p\right)\Delta^d Y_t = \left(1 + \theta_1 B + \ldots + \theta_q B^q\right)\varepsilon_t, \tag{1}$$

with the AR order p, the MA order q, and the difference order d. A three-step iterative process is used to narrow down potential candidates for the best ARIMA model to use. The steps are model identification, parameter estimation, and model diagnostics. The Box-Jenkins method does not presuppose the existence of any particular pattern in the historical data pertaining to the forecasted series. An appropriate model is sought by repeating the iterative process. The final step is to use the selected model to make predictions about future time series values.

3 Data Analysis

3.1 Stationarity

Figure 1 shows yearly passengers for the South African Airline from 1970 to 2019. The plot shows an upward trend for South African airline passengers, which increases year after year. The drop from 2007 to 2009 was caused by the global recession, but it has since picked up and trended upward. The mean is not constant in this plot, indicating that the series for South African airline passengers is not stationary. The autocorrelation function shown in Fig. 2 also demonstrates a pattern that is typical for non-stationary series. This pattern consists of a gradual decrease in the size of the autocorrelations over time. A non-stationary series can be identified by the fact that Fig. 2‘s PACF exhibits the typical characteristics of such a series, including a large spike near 1 at lag 1. This ultimately results in the first differentiation of the data.

3.1.1 Analysis for First Difference

Following the first difference, the following observations can be made from Fig. 3. The series is now indistinguishable from white noise, as there are no autocorrelations or partial autocorrelations outside of the 95% confidence intervals. This means that the annual change in the number of airline passengers is essentially a random process that is unrelated to the previous year. Tentatively we identify ARIMA (0,1,0).

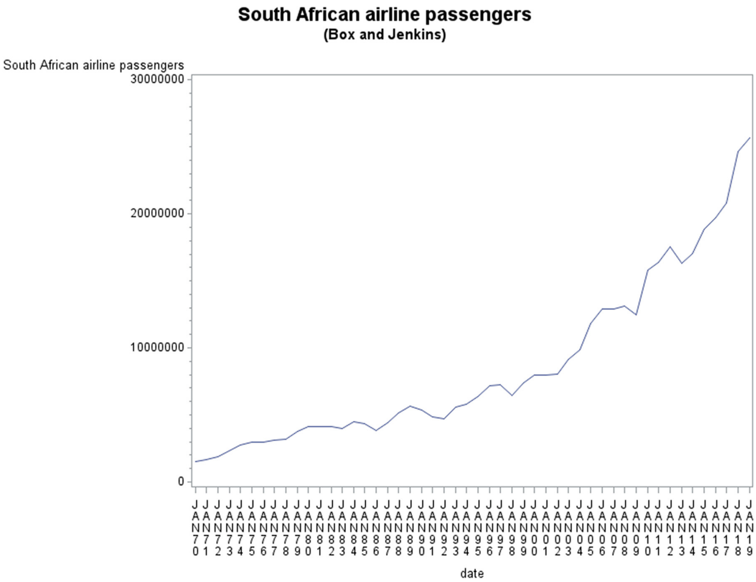


Fig. 1. South African airline passenger data

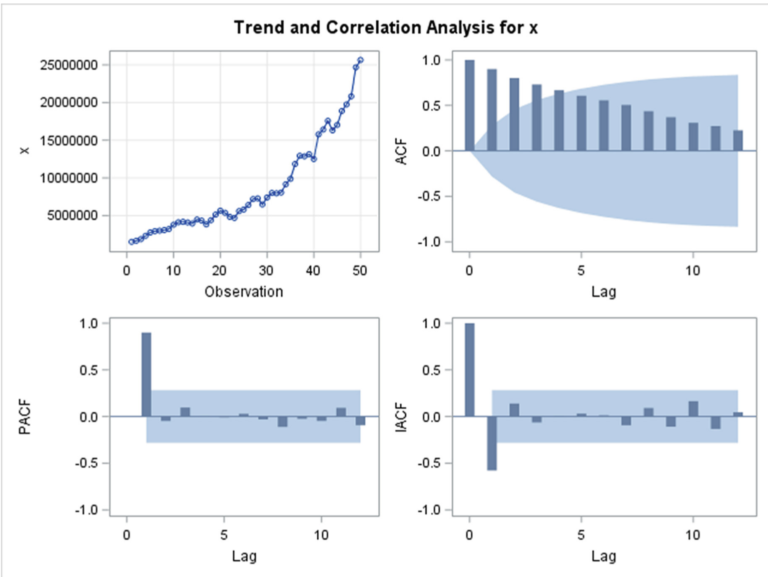


Fig. 2. Trend and correlation for the South African airline passengers

3.1.2 Identifying the Competing Models Using ESACF

Table 1 shows only one competing model, ARIMA (3,1,2) using ESACF method.

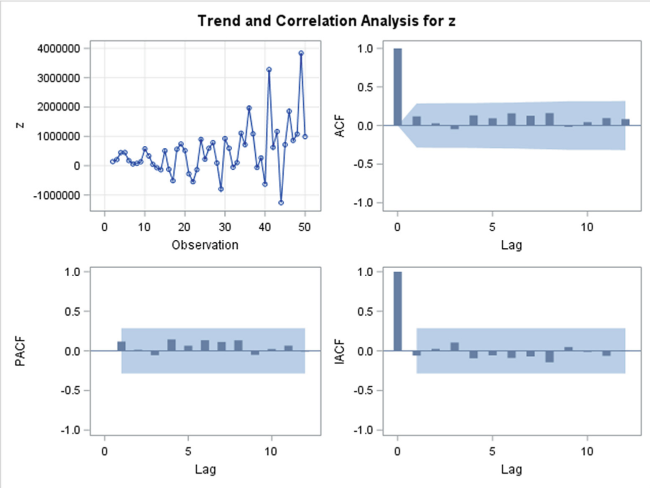


Fig. 3. Trend and correlation analysis for first differenced data

Table 1. ESACF

ARIMA ($p + d$,q) tentative order selective tests	
ESACF	
$p + d$	q
0	0
4	2

(5% significance level)

3.2 Model Estimation

3.2.1 ARIMA Model Estimates

From Table 2 we can conclude that: the MA (1), MA (2) and AR (3) terms are not statistically significant. We may need to fit ARIMA (0,1,2).

3.3 Diagnostic Checking

3.3.1 Residual Correlation Diagnosis

Figure 4 shows the residual ACF and the PACF for the model ARIMA (3,1,2). It should be noted that the ACF's residuals correlogram is flat, demonstrating that all the necessary data were gathered. The number of passengers flying on South African airlines is predicted using this model. There are no discernible lags, as shown by the flat correlogram for the residuals on the PACF.

Table 2. Maximum likelihood estimate

Maximum likelihood estimation					
Parameter	Estimate	Standard error	t value	Approx Pr > ItI	Lag
MU	494233.8	147895.7	3.34	0.0008	0
MA1,1	– 1.13123	4.12694	– 0.27	0.7840	1
MA1,2	– 1.00000	7.26741	– 0.14	0.8906	2
AR1,1	– 0.973732	0.29845	– 3.26	0.0011	1
AR1,2	– 0.78153	0.27748	– 2.82	0.0049	2
AR1,3	0.08864	0.23146	0.38	0.7017	3
Constant estimate	1317928				
Variance estimate	8.029E11				
Std error estimate	896020.5				
AIC	1489.63				
SBC	1500.981				
Number of residuals	49				

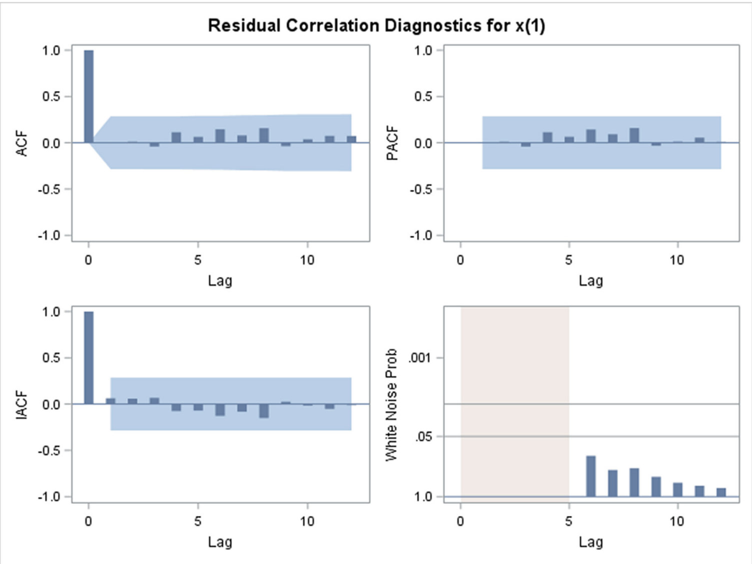


Fig. 4. Residual correlation diagnostics for ARIMA (3,1,2)

3.3.2 Residual Normality Diagnosis

The residuals' symmetrical distribution in Fig. 5 suggests that they are normally distributed. With the exception of two data points, the Quantile-Quantile (Q-Q) plot also

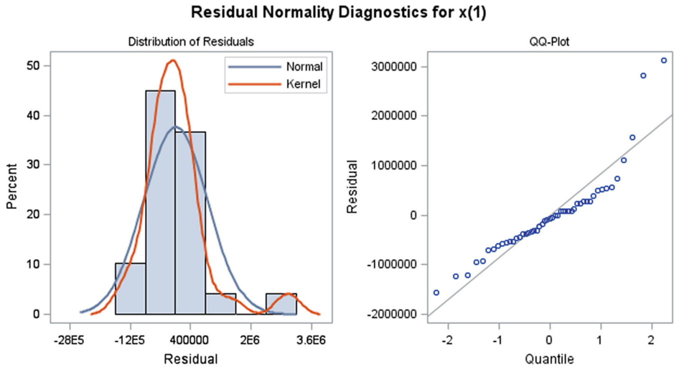


Fig. 5. Residual normality diagnostics for ARIMA (3,1,2)

demonstrates an approximately linear relationship. Therefore, the model is appropriate and usable for forecasting airline passengers in South Africa.

3.4 Forecasting

See Table 3.

Table 3. Forecasts for next five years

Autoregressive factors				
Factor 1	1 + 0.97372B**(1) + 0.78153B**(2) − 0.08864B**(3)			
Moving average factors				
Factor 1	1 + 1.13123B**(1) + 1B**(2)			
Forecasts for variable x				
Obs	Forecast	Std. Error	95% confidence limits	
51	26274540.8	896021	24518372.8	28030708.8
52	26630759.3	1370595	23944442.4	29317076.2
53	27201608.1	1754599	23762656.6	30640559.6
54	27740637.3	2023437	23774773.5	31706501.1
55	28119142.6	2284318	23641961.8	32596323.4

Figure 6 shows the forecast graph for passengers on South African airlines. Forecasts indicate a steady rise in airline travellers.

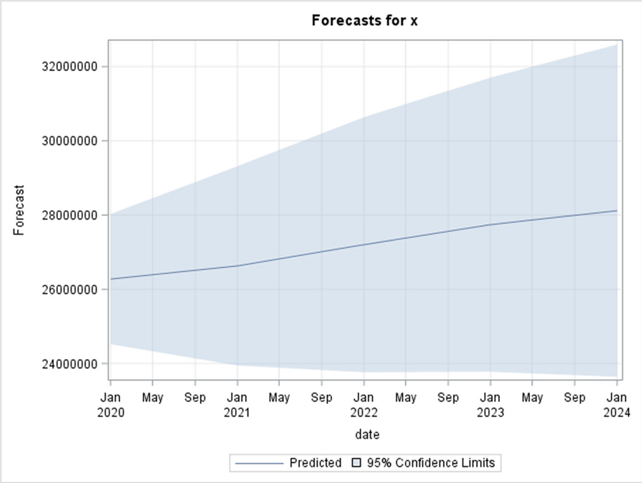


Fig. 6. Forecast plot

4 Conclusion and Recommendations

4.1 Conclusions

The ARIMA model (3,1,2) was found in this study. When the diagnostic testing was completed, this model proved its relevance and dependability. This model is adequate for forecasting airline passengers in South Africa over the next five years.

4.2 Recommendations

According to the findings of the study, it is suggested that the model be utilised on a more frequent than annual basis in order to forecast the demand for airline passenger service in South Africa., in order to make day-to-day operations run more smoothly. Many external factors can influence forecasts, both within and outside of South Africa. Changes in government, disease outbreaks, and price fixing are examples of external factors.

References

Abonazel, M.R., Abd-Elftah, A.I.: Forecasting Egyptian GDP using ARIMA models. Rep. Econ. Finan. **5**(1), 35–47 (2019)

Box, G., Jenkins, G.: Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, USA (1970)

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., Lachhab, A.: Forecasting of demand using ARIMA model. Int. J. Eng. Bus. Manag. **10**, 1–9 (2018)

Himakireeti, K., Vishnu, T.: Air passengers occupancy prediction using ARIMA model. Int. J. Appl. Eng. Res. **14**(3), 646–650 (2019)

Jadhav, V., Reddy, B.V.C., Gaddi, G.M.: Application of ARIMA model for forecasting agricultural prices. J. Agr. Sci. Tech. **19**, 981–992 (2017)

Jansson, P., Larsson, H.: ARIMA Modeling Forecasting Indices on the Stockholm Stock Exchange. Bachelor Thesis, Karlstad University (2020)

Manoj, K., Madhu, A.: An application of time series ARIMA forecasting model for predicting sugarcane production in India. In: Studies in Business and Economics, pp. 81–94 (2014)

Ming, W., Bao, Y., Hu, Z., Xiong, T.: Multistep-ahead air passengers traffic prediction with hybrid ARIMA-SVMs models. Sci. World J. **2014**(567246), 1–14 (2014)

Mishra, P., Fatih, C., Niranjan, H.K., Tiwari, S., Devi, M., Dubey, A.: Modelling and forecasting of milk production in Chhattisgarh and India. Indian J. Anim. Res. **54**(7), 912–917 (2020)

Nilgun, F., Betul, M.: Incorporating explanatory effects of neighbour airports in forecasting models for airline passenger volumes. In: Proceedings of 5th the International Conference on Operations Research and Enterprise Systems (ICORES 2016), pp. 178–185 (2016)

World Development Indicators (WDI). https://knoema.com/atlas/South-Africa/Number-of-air-passengers-carried?action=export&gadget=indicator-preview-host. Accessed 24 Nov 2021

The Sensibility of Jaya Algorithm on Tuned Mass Damper Optimization

Muhammed Çoşut, Sinan Melih Nigdeli(✉), and Gebrail Bekdaş

Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcılar, Istanbul, Turkey
{melihnig,bekdas}@iuc.edu.tr

Abstract. In this study, the optimum design of the Tuned Mass Damper (TMD), whose variables are period and damping ratio, was carried out using the Jaya algorithm by using time-domain analysis. By adding such building control systems to the structure, they contribute to keeping the displacements that may occur in the structure at certain levels, while preventing the structure from being damaged under dynamic loads. The earthquake loads affecting the structure were taken from the FEMA far-fault data as time history records and the solution to the problem was carried out with Matlab & Simulink program. The earthquake data is the data that contains the records of destructive earthquakes experienced over time and enables such systems to be designed in a controlled and effective manner. In addition, in this optimization, the maximum displacements of the structure after 10 runs by choosing different population numbers and the number of iterations, and the mean of these displacements and the standard deviation were compared. Analyzes were made for 4 different cases (50 and 100 iteration numbers—5 and 10 population numbers) and the values were recorded and it was checked how close they were to the objective function. The increase in the population number is effective in obtaining a precise optimum result.

Keywords: Tuned mass damper · Time domain optimization · Jaya algorithm · Sensibility

1 Introduction

The design of different building types has begun to be subject to changes and many limitations over time, depending on many situations. These developments update themselves depending on the natural disasters experienced, the more effective use of experimental results, and the frequency of application errors. In this way, the building systems will be designed reliably against any undesired situation. Ensuring this reliability will ensure that human life is secured by preventing fatal consequences. Capacity increases in the building can be made in many ways. In addition to being directly present in the structural carrier systems, it can be seen that in some cases they are added to the structure later and play an important role in meeting the shear forces. Friction dampers, metal dampers and viscous dampers are some of the control systems that prevent the structure

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 162–172, 2024.
https://doi.org/10.1007/978-3-031-50151-7_16

from being damaged by limiting the displacement of the structure under dynamic loads. In addition, by adding materials such as CFRP (Carbon Fibre Reinforced Polymer) to concrete, shearing capacities are increased [1, 2]. Building control systems such as TMD (Tuned Mass Damper), TLD (Tuned Liquid Damper) are placed in the structure, and these dampers, which are formed by weight, are connected by springs and provide a certain displacement of the structure under the effect of dynamic loads. In this context, many studies have been carried out. Nigdeli and Bekdaş [3] generate mass damper optimization to impede the collision of adjacent buildings under earthquake impacts. Araz and Kahya [4] used to Simulated Annealing method to find the displacement optimization of serial-tuned mass damper. Bekdaş and Niğdeli [5] used Harmony Search to reach the optimum parameter values of the TMD. Farshidianfar and Soheili [6] analyzed the TMD for high-rise buildings according to the Tabas and Kobe earthquakes using Ant Colony Optimization. By using differential evolution optimization, Valencia et al. aimed to reduce the root mean square effect of horizontal displacements and displacements at the top floor under the earthquake effect [7].

In this study, the optimization of the TMD control system in the structure according to the time domain is carried out using the Jaya algorithm. After this optimization stage, 10 different runs were made according to the number of different iterations and populations, and the maximum values of the displacements were recorded and the mean and standard deviation values were reached. In this way, the interpretations of how reaching the objective function depending on the number of iterations and populations in the optimization process were made.

2 Structure Control Systems

The movement of a building system depending on speed and acceleration under a certain force is called vibration. These vibrations are expected to be damped by the structure to prevent damage to the structure. However, in some loading situations, the structural forces will not meet and will cause undesirable situations [8, 9]. It is possible to prevent these situations by building control systems. These types of systems differ according to the type of building, the way of use, the purpose of use and the location in which they are used in the building. While some control systems are designed to be in the basement, some of them are on the middle floors and some systems are on the top floor, their designs and system features are arranged. Control systems differ depending on the material properties used. These are Active Control Systems, Passive Control Systems, Semi-Active Control Systems as well as Hybrid Control Systems.

Active Control Systems have various types which are ATMD (Active Tuned Mass Damper), Active Beam Control as well as Active Rigidity Control. There are some important advantages and disadvantages of using these control systems. Advantages of the system; it takes up less space than passive mass dampers and adjusts the movement of the system according to the magnitude of dynamic loads. Disadvantages of the system; it is expensive compared to passive dampers, vibrations are detected late and energy is supplied to the system continuously.

Passive Control Systems can be designed to have different properties by using weight, liquid and metal materials [10]. There are several Passive Control Systems such as TMD

(Tuned Mass Damper), TLD (Tuned Liquid Damper), Friction Type Dampers, Yielding Metal Dampers, Viscoelastic Dampers and Viscous Dampers. In addition, seismic isolators are used in the basement of the building, limiting the relative floor displacements of the upper floors of the building and reducing the interaction between the ground and the building. Designing it in this way will contribute to keeping the building on the safe side by providing superior behavior in some situations.

(a) (b)

Fig. 1. **a** Abernethy bridge [11], **b** Example of building [11]

(a) (b)

Fig. 2. **a** Taipei 101, **b** CN tower TV antenna

Figure 1 shows the usage of Viscous Dampers on different types of structures whereas Fig. 2 illustrates the buildings which used Tuned Mass Damper.

Semi-Active Control System was recommended in approximately 1920. This system is supported by batteries to control any loading.

Hybrid Control Systems are a type of building control system created by using active and passive control systems together. This system has been created by utilizing the power of the passive control system and the adjustment of the reaction mechanism of the active control system according to the incoming loads. The building's successful response to loadings is because it was designed with these conditions in mind [12].

Some structures which have control systems are demonstrated in Table 1.

Table 1. Control system applications

Structure	Height (m)	Location	Year	Control system
Abernethy bridge	130	Oregon, USA	1970	Viscous damper
CN tower TV antenna	553	Toronto, Canada	1973	TMD
Ando nishikicho	68	Tokyo, Japan	1993	ATMD
C office tower	130	Tokyo, Japan	1993	ATMD
Patriot stadium	–	Massachusetts, USA	2002	Viscous damper
Taipei 101	508.2	Taipei, Taiwan	2003	TMD
Shanghai world financial center	492	Shanghai, China	2008	ATMD
Canton tower (Guangzhou TV tower)	600	Guangzhou, China	2010	ATMD
Tokyo skytree	634	Tokyo, Japan	2012	TMD
432 Park avenue	426	New York, USA	2015	TMD

3 Metaheuristic Algorithm

Metaheuristic algorithms used in many fields [13] such as engineering [14], economics and transportation are algorithms inspired by natural events. It ensures that problems in most fields are solved in a short time in accordance with the standards. This has contributed to the development of the trial-and-error design process in civil engineering designs and the establishment of more effective systems. Some of the metaheuristic algorithms are Ant Colony Optimization (ACO), Teaching-Learning Based Optimization (TLBO), Bat Algorithm (BA), Jaya Algorithm, Flower Pollination Algorithm (FPA), and Differential Evolution (DE). These algorithms may differ according to the presence of control parameters and the number of phases they contain. To exemplify, TLBO has two phases while Jaya has one phase. This difference will contribute to the problems of reaching the objective function in fewer iterations.

3.1 Jaya Algorithm

Jaya algorithm [15], which means victory in Sanskrit, is one of the algorithms that always tries to reach the best. After the calculations are made according to the randomly assigned variables, the variables of the best and worst values of the objective function are found in the Jaya algorithm and used in the Eq. (1). In this way, the objective function value is reached more quickly.

$$\boldsymbol{X'_{i,\,\mathrm{new}}} = \boldsymbol{X_{i,j}} + \mathbf{rand}()\left(\boldsymbol{X_{i,\,g_{\mathrm{best}}}} - \left|\boldsymbol{X_{i,j}}\right|\right) - \mathbf{rand}()\left(\boldsymbol{X_{i,\,g_{\mathrm{worst}}}} - \left|\boldsymbol{X_{i,j}}\right|\right) \quad (1)$$

$X'_{i,\,\mathrm{new}}$ The new value of the variable.
$X_{i,j}$ The existing value of candidate solution.
$X_{i,\,g_{\mathrm{best}}}$ The best value of candidate solution.
$X_{i,\,g_{\mathrm{worst}}}$ The worst value of candidate solution.
rand() Random number.

4 Numerical Example

In order to reach the objective function of TMD optimization, structure features are given in Table 2.

Table 2. Structure properties

Symbol	Unit	Value
m: mass	kg	2924
c: damping coefficient	Ns/m	1581
k: stiffness coefficient	N/m	1390000

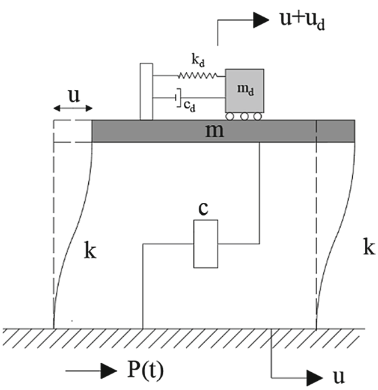


Fig. 3. TMD system

Table 3. FEMA earthquake record

	Recorded information				Motions	
Earthquake No	Year	Earthquake location	Component–1	Component–2	PGA_{max} (g)	PGV_{max} (cm/s)
1	1994	Northridge	NORTHR/MUL009	NORTHR/MUL279	0.52	63
2	1994	Northridge	NORTHR/LOS000	NORTHR/LOS270	0.48	45
3	1999	Duzce, Turkey	DUZCE/BOL000	DUZCE/BOL090	0.82	62
4	1999	Hector mine	HECTOR/HEC000	HECTOR/HEC090	0.34	42
5	1979	Imperial valley	IMPVALL/H-DLT262	IMPVALL/H-DLT352	0.35	33
6	1979	Imperial valley	IMPVALL/H-E11140	IMPVALL/H-E11230	0.38	42
7	1995	Kobe, Japan	KOBE/NIS000	KOBE/NIS090	0.51	37
8	1995	Kobe, Japan	KOBE/SHI000	KOBE/SHI090	0.24	38
9	1999	Kocaeli, Turkey	KOCAELI/DZC180	KOCAELI/DZC270	0.36	59
10	1999	Kocaeli, Turkey	KOCAELI/ARC000	KOCAELI/ARC090	0.22	40
11	1992	Landers	LANDERS/YER270	LANDERS/YER360	0.24	52
12	1992	Landers	LANDERS/CLW-LN	LANDERS/CLW-TR	0.42	42
13	1989	Loma Prieta	LOMAP/CAP000	LOMAP/CAP090	0.53	35
14	1989	Loma Prieta	LOMAP/G03000	LOMAP/G03090	0.56	45
15	1990	Manjil, Iran	MANJIL/ABBAR--L	MANJIL/ABBAR--T	0.51	54

(*continued*)

Table 3. (*continued*)

		Recorded information			Motions	
Earthquake No	Year	Earthquake location	Component–1	Component–2	PGA_{max} (g)	PGV_{max} (cm/s)
16	1987	Superstition hills	SUPERST/B-ICC000	SUPERST/B-ICC090	0.36	46
17	1987	Superstition hills	SUPERST/B-POE270	SUPERST/B-POE360	0.45	36
18	1992	Cape Mendocino	CAPEMEND/RIO270	CAPEMEND/RIO360	0.55	44
19	1999	Chi-Chi, Taiwan	CHICHI/CHY101-E	CHICHI/CHY101-N	0.44	115
20	1999	Chi-Chi, Taiwan	CHICHI/TCU045-E	CHICHI/TCU045-N	0.51	39
21	1971	San Fernando	SFERN/PEL090	SFERN/PEL180	0.21	19
22	1976	Friuli, Italy	FRIULI/A-TMZ000	FRIULI/A-TMZ270	0.35	31

By adding TMD (Fig. 3) to the established single-degree-of-freedom system, the displacements of the structure are reduced based on the FEMA (Table 3) far-fault earthquake record data. FEMA has 44 different Earthquake records which have information about locations, stations and years.

TMD design variables are shown in Table 4. While the period and damping ratio of this system, which will be optimized according to the time domain, are determined as variables, the TMD mass is determined to be 5% of the building mass. While the period value was taken between 0.5 and 1.5 times the construction period, the damping ratio was determined to be between 0.5 and 0.01. In addition, the stroke capacity of TMD is added as a constraint.

Table 4. TMD design properties and constraint

Explanation	Variables range and constraint
TMD mass	(Structure mass) $*$ 5%
TMD max. damping ratio	0.5
TMD min. damping ratio	0.01
Max. TMD period	1.5 $*$ (structure period)
Min. TMD period	0.5 $*$ (structure period)
Stroke capacity	$g(X) = \frac{\max(\lvert x_d - x_N \rvert)\mathrm{TMD}_{\mathrm{with}}}{\lvert x_N \rvert \mathrm{TMD}_{\mathrm{without}}} \leq \mathrm{st\_max}$

As a result of the analysis, the objective function was found to be 0.3789 s for the period and 0.4538 for the damping ratio. However, the analyzes were performed by repeating the Jaya algorithm 10 times according to the number of populations of 5 and 10 and both 50 and 100 iterations. As a result of some analysis, it was observed that the objective function could not be reached exactly and the displacements it created were different from each other. The finding of these different values is due to the insufficient number of iterations and the number of populations.

As a result of 10 analyzes for 4 cases, the maximum displacements were recorded separately, and the maximum and minimum displacements of these values were found, as well as the mean value and standard deviation values. The standard deviation value was calculated according to the displacement value reached by the objective function. The results of these 10 runs are shown in Table 5.

Case-1 is for 5 population numbers and 50 iterations. Case-2 is for 5 population numbers and 100 iterations. Case-3 is for 10 population numbers and 50 iterations. Case-4 is for 10 population numbers and 100 iterations.

Table 5. Case results

Explanation	Case-1	Case-2	Case-3	Case-4
Maximum displacement	0.0688	0.0693	0.0619	0.0659
Minimum displacement	0.0445	0.0424	0.0427	0.0418
Mean	0.0523	0.0544	0.0531	0.0507
Standard deviation	0.0105	0.0143	0.0106	0.00878

It can be clearly seen that all cases have different values for maximum displacement, minimum displacement, the mean of 10 runs as well as standard deviation. The maximum displacement for Case-2 is higher than others which is 0.069 m while the minimum displacement for Case-4 is lower than others. The mean of Case-4 is the closest to the objective function. Therefore, the standard deviation is found as too low compared to others. It will be seen that approximations to the objective function will increase even more after a certain iteration number is increased. Furthermore, some analyses which are related to Case-4 reached objective function. That means increasing iteration numbers and population numbers affect directly reaching objective function.

It is found that the DUZCE/BOL000 record was used for this TMD design because of its impact on the structure. Figure 4 shows the earthquake acceleration, and also Fig. 5 delineates the displacement of the structure. Moreover, Fig. 6 illustrates the total acceleration. In addition, these graphics show the change of system features both using TMD and no control system.

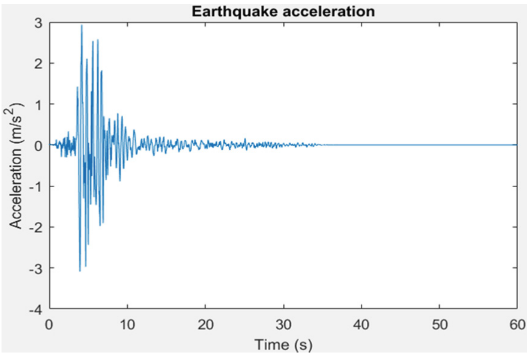


Fig. 4. Earthquake acceleration

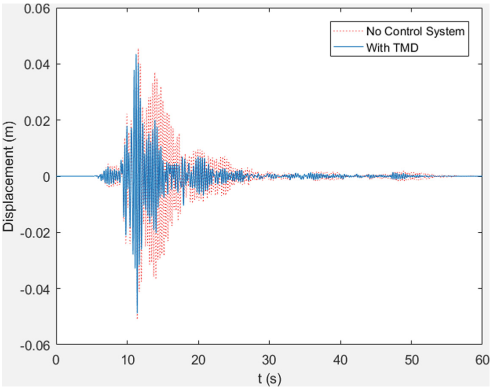


Fig. 5. Displacement of structure

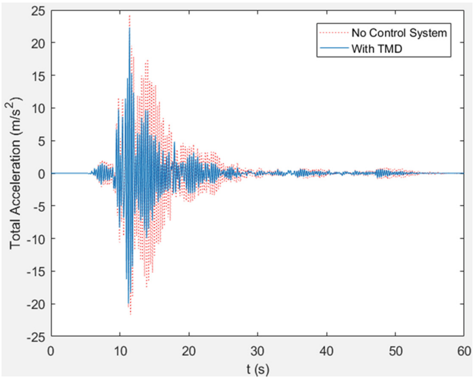


Fig. 6. Total acceleration

5 Conclusion

Building safety against dynamic loads has been applied in various ways with the development of technology and these are called building control systems. These control systems can differ as using materials, type of usage as well as the location of the systems. TMD optimization, which is one of these control systems, has been performed and the necessary properties of the system have been found according to the earthquake loads affecting the structure. As a result of this optimization, the maximum displacement value that will occur was found by using the optimum TMD damping ratio and TMD period. After these values were found, the analyses were performed 10 times separately, in 4 cases. These 4 cases were created according to the differences in the number of iterations and populations. After recording the maximum displacements as a result of 10 analyses, the maximum-minimum displacements of these states and the mean value and standard deviation values were reached. Case-4 (10 populations and 100 iterations) has the smallest minimum displacement while Case-2 (10 populations and 50 iterations) has the highest maximum displacement. Case-4 (10 populations and 100 iterations) has the smallest standard deviation value. The small standard deviation value is an indication that results closer to the displacement value found in the objective function are obtained compared to other cases. In other words, it is seen that after a certain number of iterations, the values found as a result of the optimization will approach the objective function even closer. If the number of iterations is increased and the analyses are continued, the standard deviation value will likely be much smaller. Since JA is a single-phase algorithm, the sensibility of the algorithm is improved by increasing the population number.

References

1. Çoşut, M., Bekdaş, G., Nigdeli, S.M.: Investigation of required numbers of runs for verification of Jaya algorithm for a structural engineering problem. In: Kim, J.H., et al. (ed.) Proceedings of 7th International Conference on Harmony Search, Soft Computing and Applications. Lecture Notes on Data Engineering and Communications Technologies, vol. 140, pp. 407–412. Springer, Singapore (2022)
2. Al-Ridha, A.S.D., Mahmoud, K.S., Atshan, A.F.: Effect of carbon fiber reinforced polymer (CFRP) laminates on behaviour of flexural strength of steel beams with and without end anchorage plates. Mater. Today Proc. **49**(7), 2778–2785 (2022)
3. Nigdeli, S.M., Bekdaş, G.: Optimum tuned mass damper approaches for adjacent structures. Earthq. Struct. **7**(6), 1071–1091 (2014)
4. Araz, O., Kahya, V.: Design of series tuned mass dampers for seismic control of structures using simulated annealing algorithm. Arch. Appl. Mech. **91**(10), 4343–4359 (2021). https://doi.org/10.1007/s00419-021-02013-0
5. Bekdaş, G., Niğdeli, S.M.: Estimating optimum parameters of tuned mass dampers using harmony search. Eng. Struct. **33**(9), 2716–2723 (2011)
6. Farshidianfar, A., Soheili, S.: Ant colony optimization of tuned mass dampers for earthquake oscillations of high-rise structures including soil–structure interaction. Soil Dyn. Earthq. Eng. **51**, 14–22 (2013)
7. Lara-Valencia, L.A., Caicedo-Diaz, D.A., Valencia-González, Y.: Optimal design of tuned mass dampers through differential evolution method for reducing the dynamic response of structures subjected to earthquake loads. Entramado **17**(2), 244–254. Ciencias Sociales Aplicadas (2021)
8. Montuori, R., Nastri, E., Piluso, V.: Theory of plastic mechanism control for eccentrically braced frames with inverted y-scheme. J. Constr. Steel Res. **92**, 122–135 (2014)
9. Zeynali, K., Monir, H.S., Mirzai, N.M., Hu, J.W.: Experimental and numerical investigation of lead-rubber dampers in chevron concentrically braced frames. Arch. Civ. Mech. Eng. **18**(1), 162–178 (2018)
10. Öncü-Davas, S., Alhan, C.: Reliability of semi-active seismic isolation under near-fault earthquakes. Mech. Syst. Sign. Proces. **114**, 146–164 (2019)
11. https://www.taylordevices.com/products/fluid-viscous-dampers/. Accessed 11 Feb 2023
12. Wang, L., Nagarajaiah, S., Shi, W., Zhou, Y.: Seismic performance improvement of base-isolated structures using a semi-active tuned mass damper. Eng. Struct. **271**, 114963 (2022)
13. Abdel-Basset, M., Abdel-Fatah, L., Sangaiah, A.K.: Chapter 10–Metaheuristic algorithms: a comprehensive review. In: Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications, Intelligent Data–Centric Systems, pp. 185–231 (2018)
14. Lee, K.S., Geem, Z.W.: A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice. Comput. Methods Appl. Mech. Eng. **194**(36–38), 3902–3933 (2005)
15. Rao, R.V.: Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int. J. Ind. Eng. Comput. **7**(1), 19–34 (2016)

Archimedes Optimization Algorithm on a Structural Optimization Problem

Yaren Aydin, Sinan Melih Nigdeli(✉), and Gebrail Bekdaş

Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcilar, Istanbul, Turkey

yaren.aydin1@ogr.iuc.edu.tr, {melihnig,bekdas}@iuc.edu.tr

Abstract. Optimization is required in the design of engineering problems with various design variables and constraints since the main aim of an engineer is to find the best result that provides safety and economic conditions. In this study, a cantilever beam is optimized with Archimedes Optimization Algorithm (AOA) and the results are compared with Teaching-Learning-Based Optimization and Flower Pollination Algorithm. The optimization process aims to minimize the weight of the cantilever beam with five different cross-sections. After the optimization process with these three algorithms, the results were compared. The results showed that AOA is a competitive algorithm for the design problem.

Keywords: Optimization · Metaheuristic · Cantilever beam · Flower pollination · Teaching-learning-based · Archimedes

1 Introduction

Optimization is the process carried out in order to determine the most appropriate solutions or to show the best performance for such elements while reaching the result of a problem, event or situation [1]. Optimization algorithms are divided into two categories as deterministic algorithms and stochastic algorithms. For stochastic algorithms, in general, there are two types: heuristic and metaheuristic. The heuristic is a method of trial and error when producing acceptable solutions to a complex problem in an acceptable period, but this approach is often ineffective and does not guarantee the best results. Further development over the heuristic algorithms is the metaheuristic algorithms [2]. All metaheuristic algorithms use some tradeoff between local search and global exploration. The diversity of solutions is often accomplished through randomization, which provides a good way to move from local to global-scale searching. Therefore, almost all metaheuristic algorithms are generally suitable for nonlinear modeling and general optimization [3].

Metaheuristic algorithms are effective in the optimization of structural engineering problems. Different methods have been developed and used to find the optimum design for engineering design problems. In the literature, it is seen that many studies have been done on the comparison of metaheuristic algorithms. Existing algorithms have proven their success in civil engineering problems [4–8].

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 173–180, 2024.
https://doi.org/10.1007/978-3-031-50151-7_17

Since there is no algorithm that performs best for every problem among metaheuristic methods, it remains up-to-date for researchers. Therefore, new methods are suggested. To find a better solution, many researchers have tried to propose new algorithms and improve the existing metaheuristic algorithms. Recently developed algorithms are Archimedes Optimization Algorithm (AOA) [9], Honey Badger Algorithm (HBA) [10], Prairie Dog Optimization Algorithm (PDOA) [11], etc.

In this study, existing algorithms in structural design (Teaching-Learning-based Optimization (TLBO), Flower Pollination Algorithm (FPA)) and recently developed (Archimedes Optimization Algorithm (AOA)) were employed and compared. Algorithms are briefly explained in the following sections. The unique features of the employed existing algorithms are the main reason for the selection of the algorithms. Each of these algorithms has its parameters. Changing these parameters creates differences in the local and global search abilities of the algorithm. To reach the optimum solution, three different metaheuristic algorithms were performed. According to the results, the algorithms were compared and the performances of the algorithms were evaluated.

In this study, one benchmark and its optimization on the Matlab [12] program is examined. In the benchmark, the weight of cantilever beam was minimized with three different algorithms. As a result of the research, TLBO and AOA performed the same success.

2 Methodology

2.1 Optimum Design of Cantilever Beam via Teaching-Learning-Based Optimization Algorithm

The teaching-learning-based optimization algorithm is a human-based metaheuristic algorithm developed by Rao et al. [13]. It has been very popular because characteristics of the TLBO algorithm, it does not require specific parameters, and it is fast and easy to implement [14].

TLBO has two phases as teacher phase and learner phase. In the teacher phase, the best knowledge for the whole population is considered to be the teacher, and it is used in Eq. (1). The aim is the increase the level of the average $\left(X_i^{\text{ave},t}\right)$ by the best result. A parameter called teaching factor (TF) is used and TF has a value randomly 1 or 2 (Eq. (2)). From existing solution variables $\left(X_i^{j,t}\right)$, new solution variables $\left(X_i^{j,t+1}\right)$ are generated.

$$X_i^{j,t+1} = X_i^{j,t} + \text{rand}()\left(X_{\text{best}}^{i,*} - (TF)\right)X_i^{\text{ave},t} \tag{1}$$

$$\text{TF} = \text{round}(1 + \text{rand}) \tag{2}$$

After the teacher phase, learner phase is applied. It is a simulation of a classroom in which learners increase their knowledge by interacting randomly with each other [14].

The learner phase is formulated as Eq. (3). In this process, random existing solutions are used [15].

$$x_i^{j,t+1} = \begin{cases} x_i^{j,t} + \mathrm{rand}()\left(x_i^{a,t} - x_i^{b,t}\right) \text{ if } \mathrm{f}\left(x_i^{a,t} < \mathrm{f}\left(x_i^{b,t}\right)\right) \\ x_i^{j,t} + \mathrm{rand}()\left(x_i^{b,t} - x_i^{a,t}\right) \text{ if } \mathrm{f}\left(x_i^{a,t} > \mathrm{f}\left(x_i^{b,t}\right)\right) \end{cases} \tag{3}$$

2.2 Optimum Design of Cantilever Beam via Flower Pollination Algorithm

The flower pollination algorithm is a nature-inspired metaheuristic algorithm developed by Yang [16]. The main key of the algorithm is flower constancy which is a feature about the tendency of specific pollinators to specific flowers. This feature is combined with the types of pollination and two types of optimization processes are defined. These processes are selected according to a parameter called switch probability (sp) (sp ∈ [0, 1]).

In the global pollination step that is formalized in Eq. (5), pollens travel randomly over long distances. The flight of pollinators behaves like Lévy flight behavior [17]. The parameter L is formalized in Eq. (4) and it is the strength of the pollination and a step vector. It obeys a Lévy distribution. Best existing solution $\left(X_{\mathrm{best}}^{i,*}\right)$ is 90% of flowers follow global pollination process in real life [15]. The other type is local pollination formulated as Eq. (6).

$$L = \left(\frac{1}{2\pi}\right) \times r^{-1.5} \times \mathrm{e}^{\left(\frac{1}{2r}\right)} \tag{4}$$

$$X_i^{j,t+1} = X_i^{j,t} + L\left(X_i^{j,t} - X_{\mathrm{best}}^{i,*}\right) \text{ if sp} > r_1 \tag{5}$$

$$X_i^{j,t+1} = X_i^{j,t} + \varepsilon\left(X_i^{a,t} - X_i^{b,t}\right) \text{ if sp} \leq r_1 \tag{6}$$

In local pollination, abiotic and self-pollination are used. In local pollination, a linear distribution (ε) is used as a random number between 0 and 1. Two randomly chosen existing solutions (a and b) are chosen in the modification due to self-pollination. $X_i^{a,t}$ and $X_i^{b,t}$ are random pollens from different flowers.

2.3 Optimum Design of Cantilever Beam via Archimedes Optimization Algorithm

Archimedes optimization algorithm (AOA) is a population-based metaheuristic algorithm and it is newly developed by Hashim et al. [9]. AOA is a preferred algorithm for solving optimization problems. This algorithm is based on the law of physics called Archimedes' principle. Archimedes' principle explains the law of buoyancy. Archimedes' principle states that the buoyant force is equal to the weight of the fluid displaced by the immersed object [9].

Initialize Phase

$$O_i = \mathrm{lb}_i + \mathrm{rand} \times (\mathrm{ub}_i - \mathrm{lb}_i);\; i = 1, 2, \ldots, N \tag{7}$$

O_i which presented in Eq. (7) is the random solutions of ith dimension of the ith object. ub_i and lb_i are the lower and upper bounds of the search space.

Initial density (den_i) and volume (vol_i) of each ith object are initiated randomly. Initial acceleration of each ith object (acc_i) can be defined by the same calculation as Eq. (7) as shown in Eq. (8).

$$\mathrm{acc}_i = \mathrm{lb}_i + \mathrm{rand} \times (\mathrm{ub}_i - \mathrm{lb}_i) \tag{8}$$

By using the initial solutions, the fitness function can be computed and best solution (X_{best}), density ($\mathrm{den}_{\mathrm{best}}$), volume ($\mathrm{vol}_{\mathrm{best}}$) and acceleration ($\mathrm{acc}_{\mathrm{best}}$) can be defined according to fitness function. Updating densities and volumes:

$$\mathrm{den}_i^{t+1} = \mathrm{den}_i^t + \mathrm{rand} \times \left(\mathrm{den}_{\mathrm{best}} - \mathrm{den}_i^t\right) \tag{9}$$

$$\mathrm{vol}_i^{t+1} = \mathrm{vol}_i^t + \mathrm{rand} \times \left(\mathrm{vol}_{\mathrm{best}} - \mathrm{vol}_i^t\right) \tag{10}$$

In Eqs. (9) and (10), den_i^t and vol_i^{t+1} are the density and volume of ith object at iteration $(t + 1)$.

Transfer operator (TF) refers to the transition of objects from the exploration to the exploitation phase. Here TF increases gradually over time until it reaches the current iteration number. Density decreasing factor (d) helps the algorithm on global to local search.

If $\mathrm{TF} \leq 0.5$, a collision between objects occurs (exploration phase),

$$\mathrm{acc}_i^{t+1} = \frac{\mathrm{den}_{\mathrm{mr}} + \mathrm{vol}_{\mathrm{mr}} \times \mathrm{acc}_{\mathrm{mr}}}{\mathrm{den}_i^{t+1} \times \mathrm{vol}_i^{t+1}} \tag{11}$$

The acc_i^{t+1} shown in Eq. (12) is the acceleration of the ith object at iteration $(t + 1)$. $\mathrm{acc}_{\mathrm{mr}}$, $\mathrm{den}_{\mathrm{mr}}$ and $\mathrm{vol}_{\mathrm{mr}}$ are the acceleration, density and volume of a randomly selected object.

If $\mathrm{TF} > 0.5$, there is no collision between objects (Exploitation phase). Normalize acceleration $\left(\mathrm{acc}_{i-\mathrm{norm}}^{t+1}\right)$ is the percentage of steps that each object follows in Eq. (12).

$$\mathrm{acc}_{i-\mathrm{norm}}^{t+1} = u \times \frac{\mathrm{acc}_i^{t+1} - \min(\mathrm{acc})}{\max(\mathrm{acc}) - \min(\mathrm{acc})} \tag{12}$$

Update position: If $\mathrm{TF} \leq 0.5$, the ith object's position $\left(x_i^{t+1}\right)$ shown in Eq. (13) for next iteration $t + 1$:

$$x_i^{t+1} = x_i^t + C_1 \times \mathrm{rand} \times \mathrm{acc}_{i-\mathrm{norm}}^{t+1} \times d \times \left(x_r - x_i^t\right) \tag{13}$$

If $\mathrm{TF} \leq 0.5$, the ith object's position $\left(x_i^{t+1}\right)$ shown in Eq. (14) for next iteration $t + 1$:

$$x_i^{t+1} = x_{\mathrm{best}}^t + F \times C_2 \times \mathrm{rand} \times \mathrm{acc}_{i-\mathrm{norm}}^{t+1} \times d \times \left(T \times x_{\mathrm{best}} - x_i^t\right) \tag{14}$$

In Eqs. (13) and (14), C_1, C_2, C_3 and C_4 are constants. T is TF $\times$ C_3. In Eq. (15), F is the flag to change the direction of motion and Eq. (16) is used:

$$F = \begin{cases} +1 \text{ if } P \le 0.5 \\ -1 \text{ if } P > 0.5 \end{cases} \tag{15}$$

where

$$P = 2 \times \text{rand} - C_4 \tag{16}$$

For all algorithms here, the process is finished when maximum generation is reached. If the maximum generation is not reached, all of the processes are repeated.

3 The Numerical Example

In order to evaluate the success of the optimization algorithms, the weight minimization of the cantilever beam with the design constraint is considered as an engineering problem. The beam model to be used in the cantilever beam optimization application is divided into five separate sections and has a hollow section. The l_j values in Fig. 1 represent the length of each beam section and the P value represents the vertical load applied to the beam end. The dimensions of the hollow section, which is intended to be determined so that the beam weight is minimal, are shown with x_j.

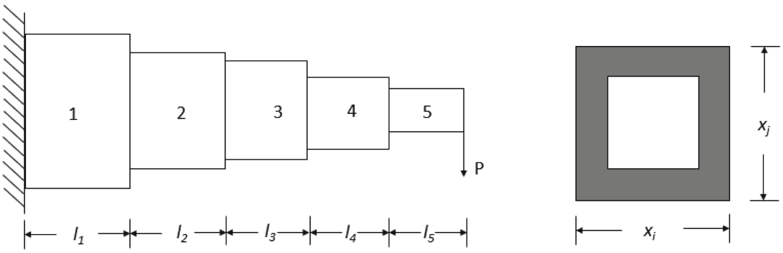


Fig. 1. Cantilever beam model and dimensions of hollow section [18]

The algorithm parameters of all algorithms are shown in Table 1. The equation expressing the minimization of the beam weight and the objective function in the optimization problem is given in Eq. (17).

$$\text{Minf}(x_j) = 0.0624(x_1 + x_2 + x_3 + x_4 + x_5) \tag{17}$$

Variation ranges for the width and depth dimensions of hollow beam section are given in Eq. (18). These are lower and upper limits design variables.

$$0.01 \le x_j \le 100 \tag{18}$$

The only constraint of the optimization problem is given in Eq. (19).

$$g_1 = \frac{61}{x_1^3} + \frac{37}{x_2^3} + \frac{19}{x_3^3} + \frac{7}{x_4^3} + \frac{1}{x_5^3} - 1 \le 0 \tag{19}$$

Table 1. Parameters of the compared algorithms

Algorithms	Parameters	Definition	Value
TLBO	p_n	Total number of learners	20
	TF	Teaching factor	1 or 2
	$t_{\max}$	Maximum iterations	10000
FPA	nf	Population/flower number	20
	sp	Key probability/search switch	0.5
	$t_{\max}$	Maximum iterations	10000
AOA	C_1	Constant	2
	C_2	Constant	6
	C_3	Constant	2
	C_4	Constant	0.5
	u and l	Range of normalization	0.9 and 0.1
	p_n	Population number	20
	$t_{\max}$	Maximum iterations	10000

4 Results

For the optimization process, it was run 20 times with 10000 iterations and optimum results were obtained. The standard deviations, mean and minimum values of the objective function and the number of iterations for which the optimum value is found for different algorithms are shown in Table 2. TLBO algorithms' number of iterations is found bigger than the other algorithms.

Table 2. Results of the Algorithms

Algorithms	Minimum value	Mean	Standard deviation	Iteration number
TLBO	1.33995638420477	1.33995642957577	2.8467961e-08	9073
FPA	1.34008063129245	1.35283946629885	0.02623517378	7372
AOA	1.33995647651313	1.33996633174065	2.5829985e-05	9050

5 Conclusions

Metaheuristic algorithms are algorithms that can give optimum solutions in a short time. Due to this feature, many metaheuristic algorithms have been proposed by researchers in the last few decades and new algorithms are still being brought to the literature. In this study, the performance of the algorithms was tested using existing and newly used algorithms in structural engineering problems.

In this study, a structural engineering problem was used to compare the performance of metaheuristic methods developed with inspiration from different sources. These algorithms are Teaching-Learning-based Optimization (TLBO), Flower Pollination Algorithm (FPA) and Archimedes Optimization Algorithm (AOA). These algorithms' characteristics were determined. The aim is to minimize the weight of the cantilever beam with five different cross-sections.

While testing the success of the algorithms, it is checked whether they find the minimum value, whether they give the same result each time, the low margin of deviation (standard deviation) between the values, and in which iteration they find the minimum value.

TLBO and AOA showed slightly the same performance in finding the minimum value. FPA follows these algorithms and obtains a minimum value very close to these algorithms.

The mean and standard deviations of the objective functions are slightly the same for TLBO and AOA. FPA is following these algorithms again.

These 3 algorithms are successful in finding the minimum value, but the number of iterations they find should also be checked. TLBO is a 2-phase algorithm, it performs double numerical operations in each iteration. In TLBO, the minimum number of computations is 9073×2. Therefore, AOA algorithm is better than TLBO algorithm. The number of iterations in which FPA reaches the optimum value is the smallest, but due to its minimum value, AOA gave better results than FPA.

Examining the overall process and results, it can be said that AOA gives better results than the other algorithms and it is competitive. In the future, AOA can be applied to additional structural engineering problems including complex ones.

References

1. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. J. Mol. Biol. **147**, 195–197 (1981)
2. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Luniver Press (2010)
3. Gandomi, A.H., Yang, X.-S., Talatahari, S., Alavi, A.H.: Metaheuristic algorithms in modeling and optimization. In: Metaheuristic Applications in Structures and Infrastructures, pp. 1–24 (2013)
4. Ocak, A., Nigdeli, S.M., Bekdaş, G., Kim, S., Geem, Z.W.: Adaptive harmony search for tuned liquid damper optimization under seismic excitation. Appl. Sci. **12**(5), 2645 (2022)
5. Bekdaş, G., Yucel, M., Nigdeli, S.M.: Evaluation of metaheuristic-based methods for optimization of truss structures via various algorithms and Lèvy flight modification. Buildings **11**(2), 49 (2021)
6. Elrehim, M.Z.A., Eid, M.A., Sayed, M.G.: Structural optimization of concrete arch bridges using Genetic Algorithms. Ain Shams Eng. J. **10**(3), 507–516 (2019)
7. Hajihassani, M., Armaghani, D.J., Kalatehjari, R.: Applications of particle swarm optimization in geotechnical engineering: a comprehensive review. Geotech. Geol. Eng. **36**(2), 705–722 (2018)
8. Mishra, M., Gunturi, V.R., Miranda, T.F.D.S.: Slope stability analysis using recent metaheuristic techniques: a comprehensive survey. SN Appl. Sci. **1**(12), 1–17 (2019). https://doi.org/10.1007/s42452-019-1707-6

9. Hashim, F.A., Hussain, K., Houssein, E.H., Mabrouk, M.S., Al-Atabany, W.: Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems. Appl. Intell. **51**(3), 1531–1551 (2021). https://doi.org/10.1007/s10489-020-01893-z
10. Hashim, F.A., Houssein, E.H., Hussain, K., Mabrouk, M.S., Al-Atabany, W.: Honey badger algorithm: new metaheuristic algorithm for solving optimization problems. Math. Comput. Simul. **192**, 84–110 (2022)
11. Ezugwu, A.E., Agushaka, J.O., Abualigah, L., Mirjalili, S., Gandomi, A.H.: Prairie dog optimization algorithm. Neural Comput. Appl. **34**(22), 20017–20065 (2022)
12. MathWorks Introduces Release 2022a of MATLAB and Simulink: Includes five new products, 11 major updates, and hundreds of new features. In: The MathWorks. Natick, Massachusetts, United States (2022). https://www.mathworks.com/company/newsroom/mathworks-introduces-release-2022a-of-matlab-and-simulink.html
13. Rao, R.V., Savsani, V.J., Vakharia, D.P.: Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput. Aided Des. **43**(3), 303–315 (2011)
14. Zou, F., Chen, D., Xu, Q.: A survey of teaching–learning-based optimization. Neurocomputing **335**, 366–383 (2019)
15. Kayabekir, A.E., Bekdaş, G., Nigdeli, S.M.: Optimum tuning of active mass dampers via metaheuristics. In: Bekdaş, G., Nigdeli, S.M. (eds.) Optimization of Tuned Mass Dampers. Studies in Systems, Decision and Control, vol. 432, pp. 155–174. Springer, Cham (2022)
16. Yang, X.S. Flower pollination algorithm for global optimization. In: Unconventional Computation and Natural Computation. UCNC 2012. Lecture Notes in Computer Science, vol. 7445, pp. 240–249. Springer, Berlin, Heidelberg (2012)
17. Kayabekir, A.E., Bekdaş, G., Nigdeli, S.M., Yang, XS.: A comprehensive review of the flower pollination algorithm for solving engineering problems. In: Nature-Inspired Algorithms and Applied Optimization. Studies in Computational Intelligence, vol. 744, pp. 171–188. Springer, Cham (2018)
18. Yang, X.-S., Bekdaş, G., Nigdeli, S.M. (eds.): Metaheuristics and Optimization in Civil Engineering. Springer International Publishing, Switzerland, Cham (2016)

A Hybrid Approach for Improving Task Scheduling Algorithm in the Cloud

Sharmin Akter[1], Md. Helal Khan[1], Lamiya Nishat[1], Farhana Alam[1], Ahmed Wasif Reza[1](✉), and Mohammad Shamsul Arefin[2](✉)

[1] Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
wasif@ewubd.edu

[2] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh
sarefin@cuet.ac.bd

Abstract. Energy consumption is a big problem for cloud computing. Different task-scheduling algorithms in cloud computing take a different amount of energy, resources, and time for executing. In this paper, our goal is to minimize the consumption of energy within a short executing time and make a high resource utilizing a task scheduling approach which will help to make sustainable green cloud computing. The common scheduling is particle swarm optimization (PSO) and ant colony optimization (ACO). This algorithm is used in the cloud. In this paper, our goal is to combine these two algorithms and make a new approach for scheduling which performance is better than the algorithms. So, our new approach is the PSAC algorithm which is a combination of PSO and ACO. PSAC approach reduces energy consumption and takes less time for execution. It also uses more resources than two algorithms. The result of the new hybrid algorithm takes less energy so energy consumption is reduced and less time for task execution. Moreover, it can utilize more resources. The result shows that the hybrid PSAC algorithm can perform the same number of tasks using lower energy and execution time than the PSO and the ACO algorithm and also utilizes more resources than the two algorithms.

Keywords: Particle swarm optimization (PSO) · PSAC · Ant colony optimization (ACO) · Green computing · Task scheduling

1 Introduction

In the modern world, a data center is vital for many companies, such as Google, Amazon, and others. However, those data centers consume a lot of energy and are expensive. That is why cloud computing is invented to reduce those problems. However, this cloud computing also consumes massive energy and produces vast environmental carbon. Massive amount of energy consumption increases the cost and affects the environment. Also, carbon is very dangerous for our environment. So, companies need to optimize the energy consumption and cost of operating. For building a green cloud data center,

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 181–193, 2024.
https://doi.org/10.1007/978-3-031-50151-7_18

we need efficient algorithms to operate it so that it reduces energy consumption and produces carbon. This paper focuses on task scheduling algorithms for cloud computing that are efficient enough, so that cloud data centers become green [1].

After analyzing the problem statements and discussing related papers, we came up with some research questions. Research questions are, Why do we improve the task scheduling algorithm for cloud computing? Why do we reduce energy consumption and make it cost-effective? Why is it necessary to reduce execution time? Why do we need to make high utilization of the cloud?

For a researcher, a major concern is to reduce the consumption of energy in cloud computing. It increases the reliability of the system. Reducing execution time and high utilization of the cloud is also a major issue for cloud computing. By improving the task scheduling algorithm, this problem can be solved. To reduce energy consumption and execution time and adequately utilize time, it is vital to improving the task scheduling algorithm. We can move into green computing and make a green data center by improving algorithms. To reduce the cost, energy consumption should reduce. Much energy is needed for a data center, which is very costly. If the data center consumes less energy, it can be reduced cost. So, the additional cost will reduce.

On the other hand, energy production harms our environment because it produces carbon. To provide service on time or quickly, we need to reduce execution time. If execution time is reduced, user can get their services within less time. A server can also give the services to more users. To provide service properly, efficiently, and timely. Proper utilization of the cloud is essential. If multi-cloud computing makes high utilization, it can give the services to the user properly, efficiently, and timely.

2 Literature Review

This section has studied related work for improving task scheduling algorithms in the cloud. In earlier works, the authors proposed a hybrid algorithm by combining other algorithms. To improve task scheduling algorithms by using a hybrid approach, we have explored related other author works.

The author combined Genetic Algorithm (GA) and Ant Colony Optimization (ACO) task scheduling algorithm. The two algorithms are combined to reduce the consumption of energy and task execution time and the improvement of the efficiency of task scheduling [2].

Sanaj and Prathap [3] proposed a chaotic squirrel search algorithm (CSSA) approach for multi-task scheduling to reduce cost and increase throughput. The algorithm provides improved coverage by organizing task allocation very well. In the end, the output of this model gives better results than the traditional SSA model.

The author explored two models: The threshold-based Task scheduling (TBTS) algorithm and the Service level agreement-based Load Balancing (SLA-LB) algorithm. TBTS algorithm schedules the task in batches. Moreover, the SLA-LB algorithm dynamically schedules the task depending on user necessities such as deadlines and budget. Those approaches reduce the time of execution and penalty cost [4].

A hybrid approach has been proposed based on combining two algorithms, Genetic Algorithm (GA) and Artificial Bee Colony (ABC). GA works for some iterations and produces the best result for the ABC algorithm. The result of GA becomes the initial input of the ABC algorithm for completing the rest of the iteration and finding the best result. This hybrid approach reduces energy consumption and makespan [5].

Another author proposed Min-Min and Max-Min algorithms for the multi-cloud network where numerous clouds linked together and provided combined service. Task scheduling is challenging for the multi-cloud comparatively single cloud. Finally, the developed model output is far more sound than the traditional models [6].

In [7], Maheswari et al. combined the Particle Swarm Optimization (PSO) algorithm and the Artificial Bee Colony (ABC) algorithm and made a new approach for reducing energy cost makespan and increasing the utilization of resources. The new combined algorithm model output is better than the two algorithms.

In [8], Pradeep and Jacob proposed a new approach using the Cuckoo search (CS) algorithm and the Harmony Search (HS) algorithm. This new approach, Cuckoo Harmony Search Algorithm (CHSA), combines those two algorithms. This approach reduces energy consumption cost, memory usage, penalty, and fitness.

In [9], Al-maamari and Omara proposed a PSOCS algorithm for task-scheduling in the cloud based on the Particle Swarm Optimization (PSO) algorithm and Cuckoo Search (CS) algorithm. The new algorithm works to reduce makespan and increase the ratio of utilization.

In [10], Uddin et al. used Ant Colony Optimization (ACO) algorithm and the Bat algorithm to make a new approach. This new approach's performance is better than the two algorithms.

In [11], Bezdan et al. used hybridized Bat algorithm for scheduling multi-object tasks in the cloud. They used the Bat algorithm only and improved it by increasing resource utilization and reducing the execution time of tasks.

In [12], Jana et al. developed a model using Particle Swarm Optimization (PSO) for the cloud environment. The model was developed in different types, such as hybrid, public, private, and community clouds. For simulation purposes, they used Cloudsim and analyzed their results with the max-min scheduling algorithm.

In [13], Zambuk et al. worked with hybrid Ant Colony Optimization (ACO) algorithm in the cloud for efficiency of energy. They used a multi-object approach to the ACO algorithm.

In [14], Barbierato et al. worked with cloud environment tools such as Cloudsim. Gridsim and Greencloud. They analyzed comparisons between them. Furthermore, make the result of scheduling algorithms without using the actual cloud.

3 Scheduling Algorithms

There are many scheduling algorithms for operating the cloud efficiently. Particle Swarm Optimization (PSO) is a scheduling algorithm used in the cloud because of its efficiency. Another efficient algorithm used in the cloud is Ant Colony Optimization (ACO). By combining those algorithms, a new model can be developed which is more efficient than those two algorithms.

3.1 Particle Swarm Optimization (PSO)

The Particle Swarm Optimization (PSO) algorithm was created by Kennedy and Eberhart in 1995. In this algorithm, several particles of a swarm move around the workspace to find the best result. For changes to the position of particles, the particles search space based on individuals' social-psychological tendency to emulate other individuals' success. The changing of the particle swarm is dependent on the neighbors. Each particle changes its current position and moves to a new one. Equation (1) is used to calculate the velocity of the particle for the next iteration, and Eq. (2) is used to calculate the position of the particle for the next iteration.

$$\begin{aligned}\text{vel}_{n+1} &= \text{vel}_n + \text{con}_1 * \text{rand}_1\left(p_{n,\,\text{best}} + \text{CP}_n\right)\\ &\quad + \text{con}_2 * \text{rand}_2\left(g_{n,\,\text{best}} + \text{CP}_n\right)\end{aligned} \tag{1}$$

$$\text{CP}_{n+1} = \text{CP}_n + \text{vel}_{n+1} \tag{2}$$

Here n refers to the number of iterations, and $n + 1$ is after the next iteration. P_{best} and g_{best} mean current best and global best experience. Con_1 and con_2 is the acceleration factor, and rand_1 and rand_2 randomly generate numbers from 0 to 1. CP is the particle's current position, and vel represents particle velocity.

Figure 1 is the flowchart of the PSO algorithm. At the first initialize the population and iteration weight. The algorithm works until it finds the best result.

3.2 Ant Colony Optimization

Marco Dorigo first introduced Ant Colony Optimization (ACO) in the 90s. A common behavior of ants is seeking a path from home to food. They reduce their distance and always use short paths for collecting the food.

This algorithm's idea is based on an ant's foraging behavior for seeking a path between its colony and source food. An ant repeatedly hops from one place to another to find food. They use pheromones so that other ants can easily reach the food using the path. In the beginning, the amount of pheromone is low. That is why it takes more time. However, as the pheromone increases, they can work quickly, and many ants join the paths. They return from the food source on some path where they spread the pheromones. All ants completed a circle tour. Like ants' behavior, the ACO algorithm works. In the cloud, tasks are scheduled by using this algorithm so that they can work quickly after some task has been completed and reduce energy (Fig. 2).

Figure 2 shows the flowchart of the ACO algorithm. In this algorithm parameter also initialize at first then it updates its pheromone and finds the best result.

4 Proposed PSAC Algorithm

The PSO algorithm has a weakness. When it increases the iteration, the rate of evolution decreases. The evaluation rate is also known as the fitness evaluation rate, which refers to the number of times the fitness function is evaluated. In the PSO algorithm, the

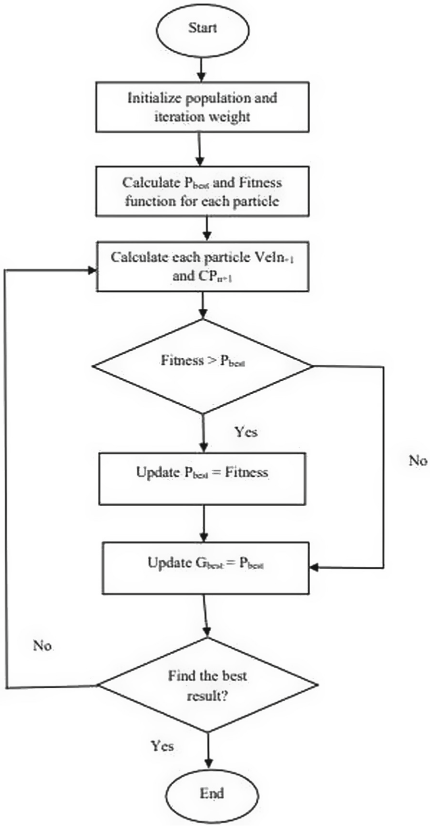


Fig. 1. Flow chart of the PSO algorithm

fitness function determines the quality of a particle's solution and is used to update the particle's velocity and position. The evaluation rate directly impacts the computational time required to find a solution that is optimal. A high evaluation rate can only lead to faster convergence to the optimal solution, but a slower convergence occurs when the evaluation rate is low. The PSO algorithm gives the best result of its first iteration. Because firstly, it has fewer particles. After every iteration, the particle increases takes more time and decreases the evolution rate.

In contrast, ACO is vice versa of PSO, which has a low rate of evolution. Because at the initial point, the pheromone of ACO is very low. However, after some time, the rate of evolution is increasing dramatically. Because after time t_a, the pheromone of ACO increase, producing the best evolution ratio.

Figure 3 demonstrates that the PSO algorithm has the highest evolution rate, whereas ACO has the lowest rate. After some iteration, both algorithms change their evolution rate. The PSO algorithm's rate of evolution is decreasing, but the evolution rate of the ACO algorithm is increasing. After some iteration or time, they both meet an optimal point. In Fig. 3, the optimal point is a. Both algorithm evolution of rate meet the point an

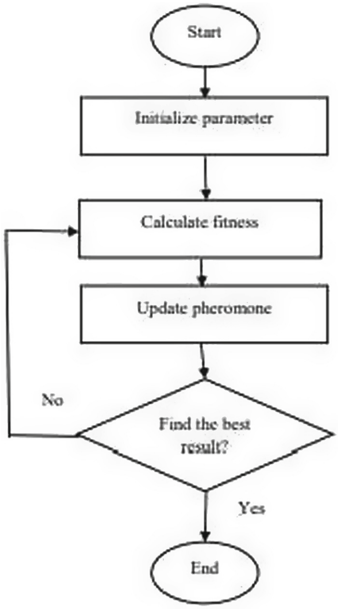


Fig. 2. Flow chart of the ACO algorithm

after time ta. Both algorithms meet a point after some time, which is the optimal solution point for switching PSO to ACO. Here the optimal point is a. Both algorithms meet this point after a time t_a.

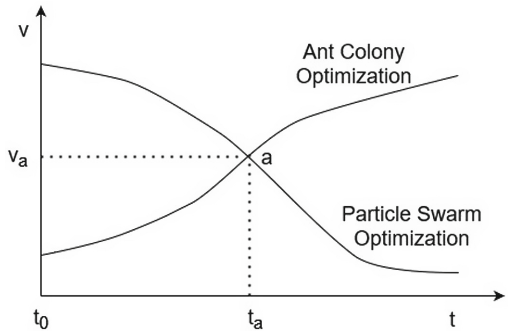


Fig. 3. Rate of evolution and time graph of PSO and ACO

In the proposed model, switching the PSO algorithm to the ACO algorithm is the optimal solution point. PSO has been worked for t_0 to t_a time. At the time of t_a the rest process will be done through the ACO algorithm. At the time of switching, the out of PSO becomes the input of ACO.

The Scheduling PSAC Algorithm

Here is the procedure for scheduling the PSAC algorithm.

1. Set the initial parameter of the PSO algorithm according to its jobs such as population size (number of the particle), mutation probability, crossover probability, and maximum and minimum iterations number.
2. The fitness function must be defined and the initial population assigned. Here the number of population is the number of particles.
3. Individually calculate the fitness value and the particle velocity. Update each particle position. And select a fitness value that is the highest. This selected fitness value will use for the next iteration.
4. The input of ACO is the out of the PSO algorithm as the optimal solution. So, the largest value of fitness of PSO becomes the pheromone in the initial state of the ACO algorithm.
5. In the PSAC algorithm, the PSO algorithm completes the first 20% of iterations. Then the largest value of fitness of PSO becomes the pheromone of the ACO algorithm. For entering the ACO, ants are randomly arranged on the nodes of resources to be allocated.
6. The next task node is selected by the ant in the ant colony. Update the pheromone on the selected node.
7. If ACO finds the best result, the PSAC algorithm terminates. And the result of the ACO algorithm is the result of the PSAC algorithm.

Figure 4 shows how our new approach works. First, initialize the PSO algorithm. After 20% of iterations of the PSO algorithm switches to the ACO algorithm. The result of PSO is the parameter of the ACO algorithm. Then the final result will find the ACO algorithm.

5 Experimental Setup

The cloudsim simulation tool is used for executing the proposed approach. This tool provides a cloud environment [15]. Table 1 shows the parameter list for the coludsim tool of this project.

Table 1 shows the different kinds of parameters used for simulation. In a cloudsim datacenter, the model works like any cloud environment datacenter hardware equipment. We take five data centers and 2–6 hosts. Here host class is related to the management of VMs. It manages memory and bandwidth and allocates the CPU of the VM. VM provides data members. Every VM has processing capacity, memory, and bandwidth. 10–180 VM are taken for the experiment. Another parameter is the task that has to be done by the algorithm. The number of tasks is 10–100 and has a length from 200–2000. Using those parameters, we simulate algorithms.

6 Result Analysis

Executing PSO, ACO, and PSAC algorithms in the Cloudsim tool to compare the algorithm's performance. We compare three performances in this paper such as time needs for execution, consumption of energy, and utilization of resources. The total number of tasks is 100, and the average task performance is 10, 30, 50, 70, and 100.

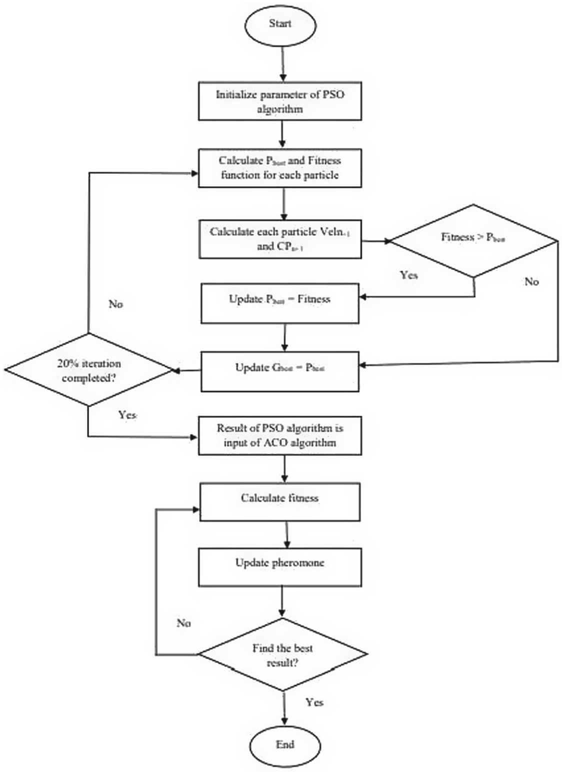


Fig. 4. Flowchart of the PSAC algorithm

Table 1. The parameter list for cloudsim

Types	Parameter	Value
Data centers	Number of data centers	5
	Number of hosts	2–6
VM	Number of VM	10–180
	Processing capacity (MIPS)	500–2000
	VM memory (RAM)	128 MB
	Bandwidth (M/s)	1000
Tasks	Number of tasks	10–100
	Task length	200–2000

Every task takes some time. Each time is different. Using the parameter of Table 1, we executed PSO, ACO, and PSAC algorithms. Then we calculated the execution time using Eq. (3).

$$\text{Execution time} = \frac{\sum(\text{start time} - \text{end time})}{\text{Total task}} \tag{3}$$

Table 2. Result for execution time in ms

Algorithm / Task number	ACO	PSO	PSAC
10	41	30	23
30	70	61	50
50	101	85	70
70	122	113	89
100	170	152	131

The graphical comparison of the result is in Fig. 5 using the value from Table 2.

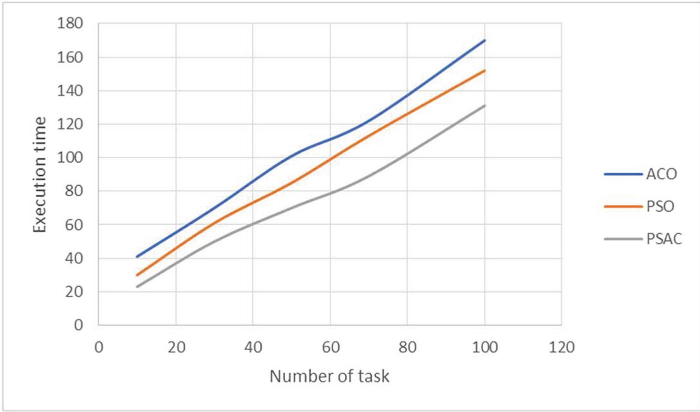


Fig. 5. Comparison between execution time and tasks

Figure 5 shows that execution time for PSO, ACO, and PSAC algorithms. PSAC algorithm takes less time than others. It can complete more tasks within a short time.

For performing a different task, there are some energy costs. Because every task needs energy for execution. We calculated the cost of energy using Eq. (4).

$$\text{Total cost} = \sum_{k=1,j=1,m=1}^{i,j,m} C_m * (T_i, \text{VM}_k) \tag{4}$$

Table 3. Result for energy consumption in mJ

Algorithm / Task number	ACO	PSO	PSAC
10	12	14	11
30	22	25	19
50	35	38	32
70	55	60	50
100	78	82	71

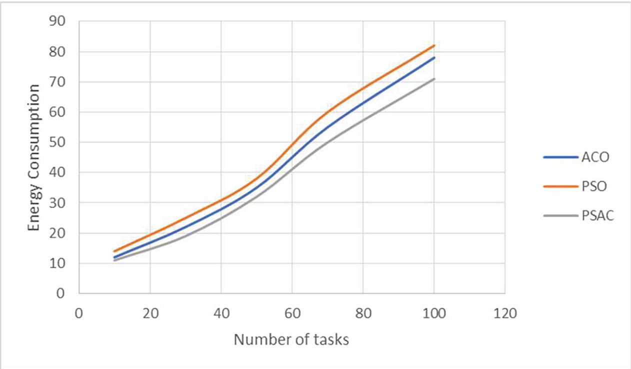


Fig. 6. Comparison between energy consumption and tasks

The graphical comparison of the result is in Fig. 6 using the value from Table 3.

Figure 6 illustrates the consumption of energy. PSAC can complete more tasks quickly than the other two algorithms. So it needs less energy to execute the tasks. PSO and ACO algorithms take more energy than PSAC. So the cost of energy of PSAC is less than two others algorithms. Moreover, when increasing the number of tasks, energy consumption decreases significantly.

The average of RAM utilization and CPU utilization calculates the resource utilization. RAM utilization is calculated by using Eq. (5).

$$\text{Utilization of RAM} = \frac{\sum \text{each task required RAM} * 100}{\text{Total RAM}} \tag{5}$$

Again, CPU utilization is calculated by using Eq. (6)

$$\text{Utilization of CPU} = \frac{\sum \text{each task required CPU} * 100}{\text{Total CPU}} \tag{6}$$

Resources utilization is the average of RAM utilization and CPU utilization.

The graphical comparison of the result is in Fig. 7 using the value from Table 4.

Table 4. Result for resource utilization in MB

Algorithm / Task number	ACO	PSO	PSAC
10	4	6	5
30	13	12	14
50	25	23	27
70	38	35	40
100	48	50	55

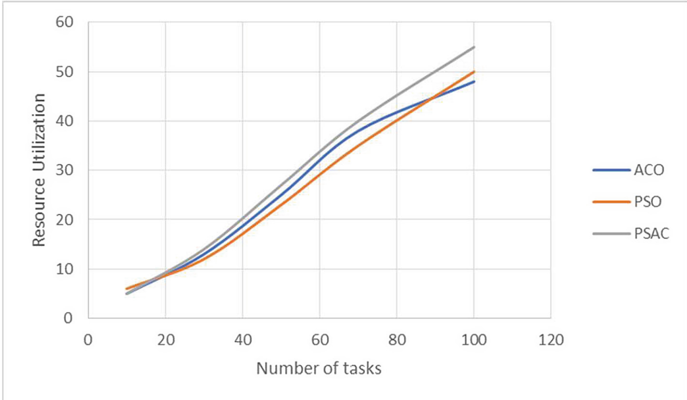


Fig. 7. Comparison between resources utilization and tasks

For resource utilization, PSAC works better than the other two algorithms. PSO and ACO give a good output when the task number is low but when the task number is huge in the cloud, PSAC utilizes more resources than the other two algorithms.

Overall, the PSAC algorithm performance is better than the two traditional algorithms. For better comparison, we calculate the time taken for execution, consumption of energy, and utilization of resources of an individual task using the value of Tables 2, 3, and 4.

Table 5. The individual task performance

Algorithm / Task number	ACO	PSO	PSAC
Execution time (ms)	2.38	1.98	1.6
Energy consumption (mJ)	0.85	0.94	0.75
Resource utilization (MB)	0.38	0.46	0.54

Table 5 clearly shows that PSAC takes less execution time and energy and utilize more resource than ACO and PSO algorithms for an individual task. In our experiment, PSAC take 78 and 38% less time for execution time for an individual task than ACO and PSO algorithm, respectively. The new approach also reduces energy consumption. For a single task, it reduces 10 and 19% energy consumption to ACO and PSO algorithms, respectively. On the other hand, the PSAC algorithm utilizes 14 and 8% more resources per task than ACO and PSO algorithms, respectively. Tables 2, 3, and 4 also demonstrate that the performance of the PSAC algorithm increases when the task increases.

7 Conclusion

PSAC is a hybrid approach of PSO and ACO algorithms. At the initial state, the PSO algorithm works better where the ACO result is poor. However, after some iteration, the result becomes vice versa. So, in the optimal solution point, switching one algorithm to another gives the best result. The combined PSAC is less energy consumption and takes less execution time than the two other algorithms. Also, the proposed approach maintains high utilization of resources. In the result analysis section, it is seen that our new approach is better than the algorithms. This approach is more efficient for cloud computing. In this paper, we work only on execution time, energy consumption, and resource utilization. This algorithm can be improved by reducing makespan, penalty cost, and complexity.

References

1. Ibrahim, H., Aburukba, R.O., El-Fakih, K.: An integer linear programming model and adaptive genetic algorithm approach to minimize energy consumption of cloud computing data centers. Comput. Electr. Eng. **67**, 551–565 (2018)
2. Zong, Z.: An improvement of task scheduling algorithms for green cloud computing. In: 2020 15th International Conference on Computer Science & Education (ICCSE), pp. 654–657. IEEE (2020)
3. Sanaj, M.S., Prathap, P.M.J.: Nature inspired chaotic squirrel search algorithm (CSSA) for multi objective task scheduling in an IAAS cloud computing atmosphere. Eng. Sci. Technol. Int. J. **23**(4), 891–902 (2020)
4. Lavanya, M., Shanthi, B., Saravanan, S.: Multi objective task scheduling algorithm based on SLA and processing time suitable for cloud environment. Comput. Commun. **151**, 183–195 (2020)
5. Kumar, S., Kalra, M.: A hybrid approach for energy-efficient task scheduling in cloud. In: Proceedings of 2nd International Conference on Communication, Computing and Networking. Lecture Notes in Networks and Systems, vol. 46, pp. 1011–1019. Springer, Singapore (2019)
6. Panda, S.K., Gupta, I., Jana, P.K.: Task scheduling algorithms for multi-cloud systems: allocation-aware approach. Inf. Syst. Front. **21**(2), 241–259 (2019). https://doi.org/10.1007/s10796-017-9742-6
7. Maheswari, P.U., Edwin, E.B., Thanka, M.R.: A hybrid algorithm for efficient task scheduling in cloud computing environment. Int. J. Reasoning Based Intell. Syst. **11**(2), 134–140 (2019). https://doi.org/10.1504/IJRIS.2019.10021325

8. Pradeep, K., Jacob, T.P.: A hybrid approach for task scheduling using the cuckoo and harmony search in cloud computing environment. Wireless Pers. Commun. **101**(4), 2287–2311 (2018). https://doi.org/10.1007/s11277-018-5816-0
9. Al-maamari, A., Omara, F.A.: Task scheduling using hybrid algorithm in cloud computing environments. IOSR J. Comput. Eng. (IOSR-JCE) **17**(3, Ver. VI), 96–106 (2015)
10. Uddin, M.Y., et al.: Development of a hybrid algorithm for efficient task scheduling in cloud computing environment using artificial intelligence. Int. J. Comput. Commun. Control **16**(5), 1–12 (2021)
11. Bezdan, T., Zivkovic, M., Bačanin, N., Strumberger, I., Tuba, E., Tuba, M.: Multi-objective task scheduling in cloud computing environment by hybridized bat algorithm. J. Intell. Fuzzy Syst. **42**(1), 411–423 (2022)
12. Jana, B., Chakraborty, M., Mandal, T.: A task scheduling technique based on particle swarm optimization algorithm in cloud environment. In: Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol. 742, pp. 525–536. Springer, Singapore (2019)
13. Zambuk, F.U., Ya'u Gital, A., Jiya, M., Gari, N.A.S., Ja'afaru, B., Muhammad, A.: Efficient task scheduling in cloud computing using multi-objective hybrid ant colony optimization algorithm for energy efficiency. Int. J. Adv. Comput. Sci. Appl. **12**(3), 450–456 (2021)
14. Barbierato, E., Gribaudo, M., Iacono, M., Jakóbik, A.: Exploiting CloudSim in a multiformalism modeling approach for cloud based systems. Simul. Model. Pract. Theory **93**, 133–147 (2019)
15. Mosa, A., Sakellariou, R.: Virtual machine consolidation for cloud data centers using parameter-based adaptive allocation. In: ECBS'17 Proceedings of the Fifth European Conference on the Engineering of Computer-Based Systems, Article No. 16, pp. 1–10 (2017)

Design Optimization of Tuned Liquid Dampers with Hybrid Algorithms

Ayla Ocak(✉), Sinan Melih Nigdeli, and Gebrail Bekdaş

Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcılar, Istanbul, Turkey

aylaocak@outlook.com, {melihnig,bekdas}@iuc.edu.tr

Abstract. Tuned liquid damping devices are a low-cost, passive control system that does not require an external energy source and is effective in damping the dynamic vibrations of structures. Tuning the damper tank geometry and device parameters is important in optimizing the efficiency to be obtained from the control. In this study, tuned liquid damper (TLD) device parameters were optimized with algorithms created by hybridizing metaheuristic algorithms to minimize the structure movement. For this purpose, Jaya (JA) and Teaching-Learning Based Optimization (TLBO) algorithms were hybridized. By optimizing the properties of the TLD, the maximum displacement and total acceleration values of the structure under the critical earthquake record were reduced. In light of the results obtained, it was determined that hybrids may increase the performance of the algorithm in some cases.

Keywords: Tuned liquid damper · Passive control · Metaheuristic algorithm · Hybrid algorithm · Jaya algorithm · Teaching-learning-based optimization

1 Introduction

Tuned liquid damping devices (TLDs) are control systems that are attached to the structure with a spring and absorb the energy of the system by oscillating a liquid in the opposite direction to the movement of the structure by using the kinematic viscosity of the liquid mass inside. The basic device parameters that affect damping are the geometric properties of the TLD tank such as liquid height and floor length, and the period and damping rate of the TLD. Fujino et al. suggested in their study that the ratio of the liquid height to the effective base length of the TLD device should not exceed 15% [1]. In another study, it was determined that increasing the TLD tank bottom length for constant liquid height decreased TLD performance and that TLD tank geometry designs should be directed toward longer designs that are oriented rather vertically than horizontally [2]. It has been observed that TLD devices with flat bottoms are more robust than curved ones [3, 4]. According to the linear hydrodynamic theory, some of the liquid in the liquid dampers does not participate in the sloshing but acts as a solid mass with the tank [5]. The damper period and damping rate are properties that depend on the sloshing liquid mass. In determining these properties, optimization is required to maximize the efficiency of the damper device.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 194–204, 2024.
https://doi.org/10.1007/978-3-031-50151-7_19

Optimization is the process of idealizing the desired variable within the determined objective function. Various metaheuristic algorithms have been produced as optimization tools with the development of mathematical models derived by imitating nature and living life [5–11]. There are damper optimization studies that support the effectiveness of metaheuristic algorithms [12–17]. Classical tuned liquid dampers (TLDs), TLDs with different viscosity liquids, magneto-rheological (MR) liquid TLDs, tuned liquid column dampers (TLCDs), and multi-tuned liquid column inerter dampers (MTLCDI) have been optimized in various studies [18–22]. Optimization studies were done with metaheuristic algorithms such as genetic algorithm, firefly algorithm, harmony search algorithm, and particle swarm optimization [18–22]. The use of hybrid algorithms can produce practical solutions in minimizing the negative effects of the unique design factors of the algorithms on optimization. Amini and Ghaderi obtained a hybrid of ant colony optimization with the harmony search algorithm and obtained more effective results than HS in the problem of optimal placement of the damper [23]. Miguel et al. hybridized the firefly algorithm with the Nelder mead algorithm and found that the computational cost for multi-tuned mass dampers (MTMDs) outperformed the optimization with the firefly algorithm [24]. Nigdeli et al. used a hybrid of flower pollination algorithm and harmony search algorithm in the optimization of a tuned mass damper (TMD) placed on the top of a structure exposed to earthquake excitations [25]. Nigdeli et al. provided a more effective convergence with the hybrid algorithm compared to the classical flower pollination algorithm in reaching the optimum result.

In this study, a passively tuned liquid damping device was placed on a single-storey structure and optimized with a hybrid metaheuristic algorithm. In the optimization, two hybrid algorithms created from Jaya algorithm (JA) and teaching-learning-based optimization (TLBO) were used. Various TLD parameters are optimized to minimize structure movement under earthquake excitation. A model was created with 3 different liquids as TLD tank liquid. Earthquake analysis was performed using the optimum parameters obtained by the optimization with hybrid algorithms for the earthquake record that is critical for the sample structure model.

2 Methodology

In this section, basic motion equations of passively tuned liquid dampers, calculations of device parameters and algorithm equations of hybrid algorithms to be used in optimization are given.

2.1 TLD Parameters and Equations of Motion

TLD devices benefit from the sloshing mechanism of the liquid and it provides structural control. It is known that the liquid in the TLD tank is not sloshing completely but is a passive liquid mass that behaves like a solid mass by acting in the same way as the tank. Since not all of the liquid is shaken, the liquid is separated into an active and passive liquid mass. The calculation of the sloshing active liquid mass is given in Eq. (1), and the damper mass acting as a solid mass, which is the sum of the passive liquid mass and the mass of the tank, is given in Eq. (2) [18, 26].

$$m_s = m_{\mathrm{st}} \times R \times \frac{\tanh\left(\frac{1.84h}{R}\right)}{2.2h} \tag{1}$$

$$m_d = m_{\mathrm{TLD}} - m_s \tag{2}$$

In the equations, h is the liquid height in the tank, R is the bottom radius of the cylindrical tank, m_{st} is the total liquid mass, m_s is the sloshing liquid mass, m_d is the passive liquid acting as a solid mass and the mass of the tank and m_{TLD} is the total TLD mass. The terms with index *"d"* in the equations refer to the damper, and the terms with index *"s"* refer to the terms of the sloshing liquid. Damper and sloshing liquid stiffnesses (k_d, k_s) are given in Eqs. (3) and (4), respectively, and damping coefficients (c_d, c_s) are given in Eqs. 5 and 6 [18, 26].

$$k_d = m_d \times \left(\frac{2\pi}{T_d}\right)^2 \tag{3}$$

$$k_s = m_{\mathrm{st}} \times \frac{g\left\{\tanh\left(\frac{1.84h}{R}\right)\right\}^2}{1.19h} \tag{4}$$

$$c_d = 2 \times \zeta_d \times \sqrt{m_d \times k_d} \tag{5}$$

$$c_s = \zeta_s \times 2\sqrt{m_s k_s} \tag{6}$$

The damper period T_d, the damper damping ratio, is represented by the damping ratio of the sloshing liquid. The damping ratio equation is shown ζ_d in Eq. (7) and the damping ratio equation of sloshing liquid is shown ζ_s in Eq. (8) [18, 26].

$$\zeta_d = \frac{c_d}{2m_d\sqrt{\frac{k_d}{m_d}}} \tag{7}$$

$$\zeta_s = 4.98v^{\frac{1}{2}}R^{-\frac{3}{4}}g^{-\frac{1}{4}}\left[1 + \frac{0.318}{\sinh\left(\frac{1.84h}{R}\right)}\frac{1 - \frac{h}{R}}{\cosh\left(\frac{1.84h}{R}\right)}\right] \tag{8}$$

In the equations, the kinematic viscosity of the liquid is denoted by v and the gravitational acceleration g. The basic equation of motion of a system consisting of the motion of the sloshing liquid, the TLD tank, and the structure is shown in Eq. (9).

$$[M]\{\ddot{X}\} + [C]\{\dot{X}\} + [K]\{X\} = -[M]1\{\ddot{X}_g\} \tag{9}$$

The movement of the TLD structure under earthquake excitation is expressed X, by its speed $\dot{X}$, acceleration $\ddot{X}$, and ground acceleration $\ddot{X}_g$. M, K, and C are the mass, stiffness, and damping coefficient matrices of the system, respectively.

2.2 Optimization with Hybrid Metaheuristics Algorithms

Metaheuristic algorithms are optimization methods inspired by natural events and living behavior. Rao et al., inspired by the learning and teaching process of a class, group, or community, and the knowledge transfer process within them, developed two-phase teaching-learning-based optimization [27]. Here, in the first phase, the teacher is training a group, while in the second phase, the interaction between the students is mentioned. The optimum solution equations for these two stages are given in Eqs. (10) and (11).

$$X_{\text{new}} = X_{\text{old}} + \text{rand}(0,\ 1)\left(g^* - T_F X_{\text{mean}}\right) \tag{10}$$

$$X_{\text{new}} = \begin{cases} X_{\text{old}} + \text{rand}(0,\ 1)\left(X_j - X_k\right);\ \text{f}\left(X_j\right) > \text{f}(X_k) \\ X_{\text{old}} + \text{rand}(0,\ 1)\left(X_k - X_j\right);\ \text{f}(X_k) > \text{f}\left(X_j\right) \end{cases} \tag{11}$$

T_F in the equations represents the teaching factor varying between 0 and 1, X_{mean} is the mean of the population, X_{new} is the new solution, X_{old} is the old solution, g^* is the best solution in the solution vector, X_j and X_k are two students selected from the educated class.

By reducing the two-phase teaching-learning process to a single phase, the Jaya algorithm was produced [28]. The solution reduced to a single equation accelerated the optimization of the production process. The solution generation equation of the Jaya algorithm is shown in Eq. (12).

$$\begin{aligned} X_{\text{new}} = & X_{\text{old}} + \text{rand}(0,\ 1)\left(g^* - |X_{\text{old}}|\right) \\ & - \text{rand}(0,\ 1)\left(g^w - |X_{\text{old}}|\right) \end{aligned} \tag{12}$$

The worst solution in the solution vector is represented by g^w.

The JAS is a hybrid algorithm in which there is a student phase (Eq. (11)) and Jaya as a second phase. Based on a randomly assigned value, a phase is chosen in an iteration. A solution is produced with the knowledge-sharing process of the TLBO algorithm, which is made by the students after the learning, or with Jaya. In the hybrid algorithm called JA2S, the production process in Eqs. (11) and (12) is consequently used in an iteration.

The main objective function in the optimization of damping devices is to minimize displacement. Controlling the displacement is the main objective function for optimizing the dampers used in vibration control. The objective function of the dampers to reduce the displacement is shown in Eq. (13).

$$\text{f}(x) = \max(|X|) \tag{13}$$

3 The Numerical Example

In this study, a three-degree-of-freedom (3DOF) structure model was created by placing a two-degree-of-freedom liquid damper on top of a single-degree-of-freedom structure. In the damper tank placed in the structure, water, oil, and glycerin liquids were used separately in the study. The behavior of the controlled structure under FEMA P-695 far-fault earthquake records was investigated by the simulation created via Matlab Simulink to the structure [29, 30]. The mass of the structural model used in the study is 100 t, the stiffness is 3.95 MN/m, the damping coefficient is 0.06 MNs/m, the period is 1 s, and the structure mass ratio is 5% for TLD. The mechanical model of the TLD structure is shown in Fig. 1.

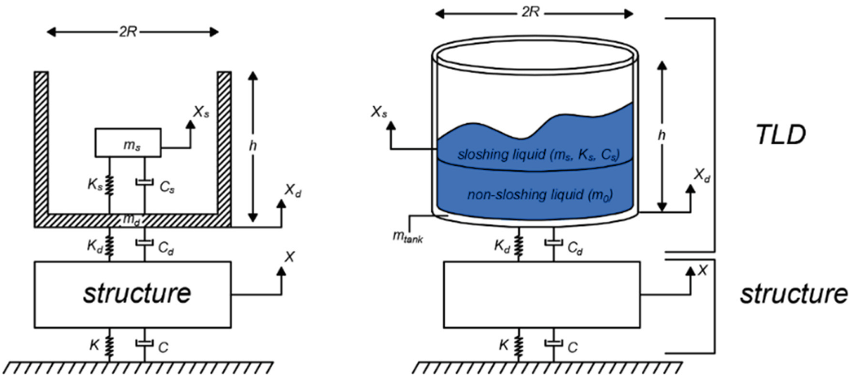


Fig. 1. Structure + TLD system model

Liquid height in the tank, tank radius, damper period, and damping ratio parameters affecting damping have been optimized by hybrid algorithms. The design constraints used in the optimization are given in Table 1. In the optimization, the population number is 10 and the iteration number is 1000. The optimum TLD parameters obtained as a result of the optimization are given in Table 2 for the TLD structure model containing 3 different liquids.

Table 1. TLD design constraints

Explanation	Design parameter
Minimum-maximum height	0.1–10.0 m
Minimum-maximum radius	0.1–10.0 m
TLD period	0.5–1.5 s of the structure period
Height-to-radius ratio	$\frac{h}{2R} > 0.15$
Damping ratio	1–50%

Structure displacement and total acceleration values obtained from earthquake analysis, which is critical for the structure with optimum TLD parameters, are shown in Table 3. The time-dependent graphs of the structural displacement and total acceleration obtained in the critical earthquake recording are given in Figs. 2, 3, 4 for the JAS hybrid and in Figs. 5, 6, 7 for the JA2S hybrid in TLD-water, TLD-oil and TLD-glycerine system models.

Table 2. Optimum results for TLD

Variables	Optimized values					
	TLD-Water		TLD-Oil		TLD-Glycerine	
	JAS	JA2S	JAS	JA2S	JAS	JA2S
T_d (s)	0.8910	0.8906	0.8861	0.8862	0.8828	0.8852
ζ_d	0.0260	0.0256	0.0202	0.0203	0.0100	0.0100
R (m)	0.5575	0.5575	0.6013	0.6012	0.6217	0.5680
h (m)	2.2584	4.5320	3.2849	3.3455	2.8105	3.7154

4 Discussion and Conclusions

In this study, the effects of tuned liquid dampers in three different liquids were investigated by optimizing with hybrid algorithms created with one and two phases in an iteration. Earthquake analysis was performed on a three-degree-of-freedom TLD model, with a structured period of 1 s. By increasing the single phase in the Jaya algorithm to two, the optimization performance of the hybrid algorithm was investigated. The results obtained with TLD optimization can be summarized as follows:

1. When all displacement values were examined, optimum values were obtained for all liquids in the JA2S hybrid where the phase was increased to 2 in an iteration. There was a performance difference not exceeding 1% between them.
2. As the results were close to each other for the three liquids, it was observed that the optimum liquid was glycerine with a displacement reduction of 19.73%.
3. Looking at the optimized values in Table 2, it was seen that there were different values in the h and R values of TLD tank parameters in the 2 hybrid algorithms. Considering these results, it is understood that more than one tank geometry can be created for optimum displacement values and design alternatives will gain diversity with optimization.
4. The classical form of JA and TLBO have also comparative and similar performances in the reduction of the displacement, but the glycerine case has slightly better results for hybrids.

In the analysis of tuned liquid dampers, it was determined that the optimization offers different design alternatives and the changes to be made in the phase part of the hybrid algorithms can affect the efficiency in several cases. It has been observed that the optimization and hybridization to be made in the algorithm parameters in the optimization with metaheuristic algorithms will positively affect the optimization process and it is an area open to research.

Table 3. Displacement and total acceleration values obtained from critical earthquake analysis with TLD optimization

Algorithm	Without TLD structure		TLD-Water		TLD-Oil		TLD-Glycerine	
	Displacement (m)	Total acceleration (m/s^2)	Displacement (m)	Total acceleration (m/s^2)	Displacement (m)	Total acceleration (m/s^2)	Displacement (m)	Total acceleration (m/s^2)
JAS	0.2873	11.4078	0.2315	8.7603	0.2322	8.7625	0.2323	8.8538
JA2S			0.2315	8.7575	0.2322	8.7630	0.2306	8.7244
JA [31]			0.2315	8.7560	0.2322	8.7631	0.2324	8.8551
TLBO [31]			0.2315	8.7553	0.2322	8.7626	0.2324	8.8554

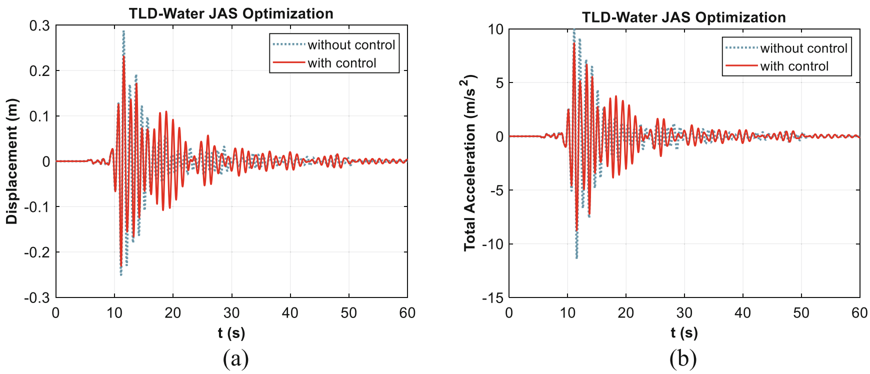


Fig. 2. Time-dependent graphs for critical earthquake record analysis of the JAS hybrid-optimized TLD-Water: **a** Structure displacement, **b** Structure total acceleration

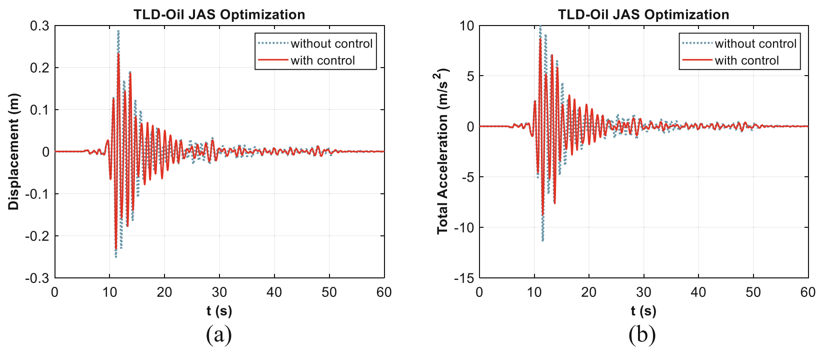


Fig. 3. Time-dependent graphs for critical earthquake record analysis of the JAS hybrid-optimized TLD-Oil: **a** Structure displacement, **b** Structure total acceleration

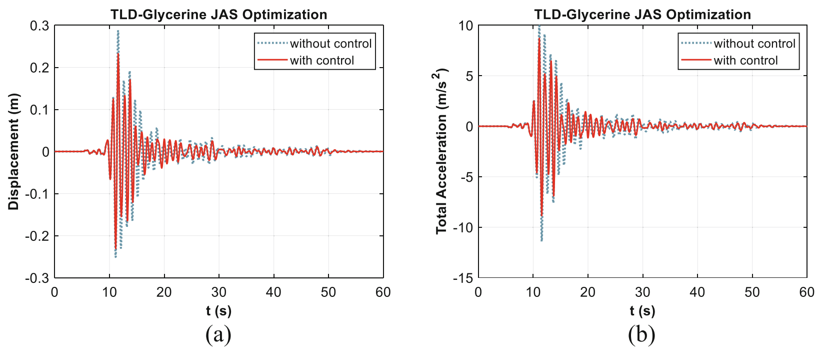
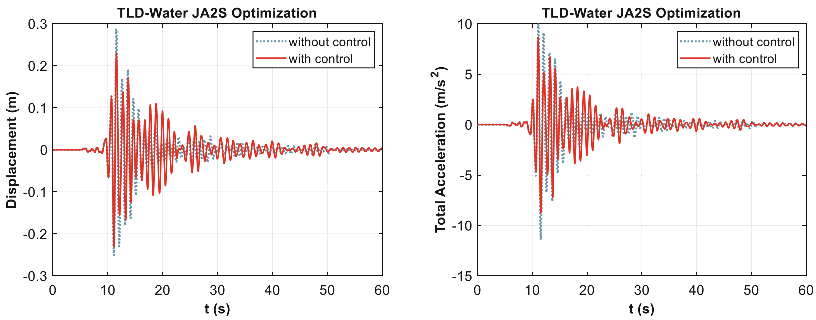
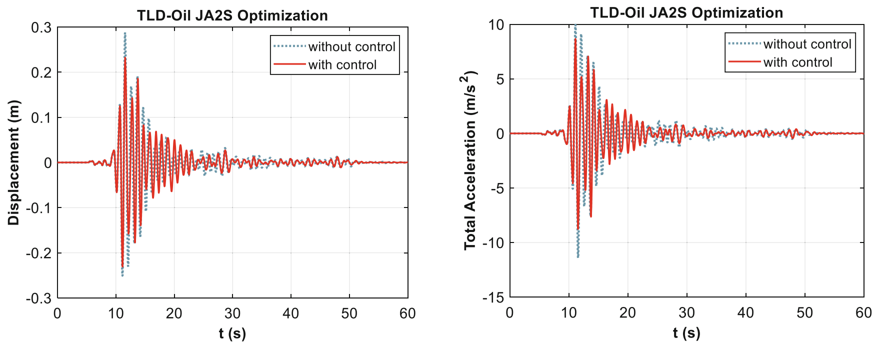


Fig. 4. Time-dependent graphs for critical earthquake record analysis of the JAS hybrid-optimized TLD-Glycerine: **a** Structure displacement, **b** Structure total acceleration



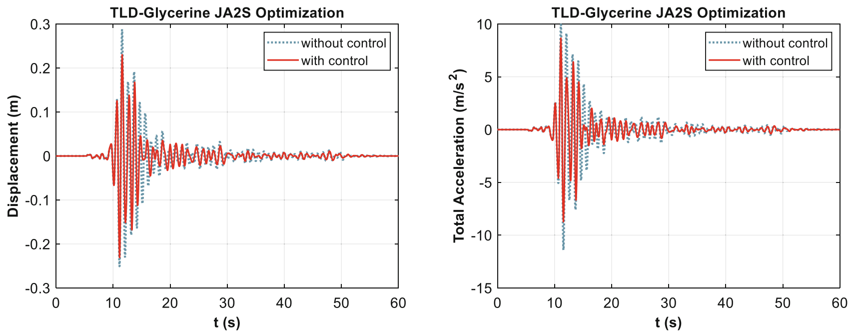
(a) (b)

Fig. 5. Time-dependent graphs for critical earthquake record analysis of the JA2S hybrid-optimized TLD-Water: **a** Structure displacement, **b** Structure total acceleration



(a) (b)

Fig. 6. Time-dependent graphs for critical earthquake record analysis of the JA2S hybrid-optimized TLD-Oil: **a** Structure displacement, **b** Structure total acceleration



(a) (b)

Fig. 7. Time-dependent graphs for critical earthquake record analysis of the JA2S hybrid-optimized TLD-Glycerine: **a** Structure displacement, **b** Structure total acceleration

References

1. Fujino, Y., Sun, L., Pacheco, B.M., Chaiseri, P.: Tuned liquid damper (TLD) for suppressing horizontal motion of structures. J. Eng. Mech. **118**(10), 2017–2030 (1992)
2. Ocak, A., Nigdeli, S.M., Bekdaş, G.: Investigation of effective liquid length on structure movement in tuned liquid dampers. In: 26th International Conference “MECHANIKA 2022”, Lithuania (2022)
3. Zhang, Z.: Numerical and experimental investigations of the sloshing modal properties of sloped-bottom tuned liquid dampers for structural vibration control. Eng. Struct. **204**, 110042 (2020)
4. Cavalagli, N., Agresta, A., Biscarini, C., Ubertini, F., Ubertini, S.: Enhanced energy dissipation through 3D printed bottom geometry in Tuned Sloshing Dampers. J. Fluids Struct. **106**, 103377 (2021)
5. Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI (1975)
6. Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P.: Optimization by simulated annealing. Science 220(4598), 671–680 (1983)
7. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: an autocatalytic optimizing process. Technical Report 91-016, pp. 1–21 (1991)
8. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN’95—International Conference on Neural Networks, vol. 4, pp. 1942–1948. IEEE (1995)
9. Geem, Z.W., Kim, J.H., Loganathan, G.V.: A new heuristic optimization algorithm: harmony search. SIMULATION **76**(2), 60–68 (2001)
10. Yang, X.S.: Flower pollination algorithm for global optimization. In: Unconventional Computation and Natural Computation. Lecture Notes in Computer Science, vol. 7445, pp. 240–249 (2012)
11. Karaboğa, D.: An Idea Based on Honey Bee Swarm for Numerical Optimization. Technical Report-TR06, pp. 1–10. Erciyes University, Engineering Faculty, Computer Engineering Department (2005)
12. Bekdaş, G., Nigdeli, S.M.: Mass ratio factor for optimum tuned mass damper strategies. Int. J. Mech. Sci. **71**, 68–84 (2013)
13. Nigdeli, S.M., Bekdaş, G.: Optimum tuned mass damper design in the frequency domain for structures. KSCE J. Civ. Eng. **21**(3), 912–922 (2017)
14. Siami, A., Karimi, H.R., Cigada, A., Zappa, E., Sabbioni, E.: Parameter optimization of an inerter-based isolator for passive vibration control of Michelangelo’s Rondanini Pietà. Mech. Syst. Signal Process. **98**, 667–683 (2018)
15. Yang, Y., Li, C.: Performance of tuned tandem mass dampers for structures under the ground acceleration. Struct. Control. Health Monit. **24**(10), e1974 (2017)
16. Djerouni, S., Abdeddaim, M., Elias, S., Rupakhety, R.: Optimum double mass tuned damper inerter for control of structure subjected to ground motions. J. Build. Eng. **44**, 103259 (2021)
17. Raeesi, F., Azar, B.F., Veladi, H., Talatahari, S.: An inverse TSK model of MR damper for vibration control of nonlinear structures using an improved grasshopper optimization algorithm. Structures **26**, 406–416. Elsevier (2020)
18. Ocak, A., Bekdaş, G., Nigdeli, S.M., Kim, S., Geem, Z.W.: Optimization of tuned liquid damper including different liquids for lateral displacement control of single and multi-story structures. Buildings **12**(3), 377 (2022)
19. Talib, M.H.A., Darus, I.Z.M.: Intelligent fuzzy logic with firefly algorithm and particle swarm optimization for semi-active suspension system using a magneto-rheological damper. J. Vib. Control **23**(3), 501–514 (2017)

20. Park, S., Glade, M., Lackner, M.A.: Multi-objective optimization of orthogonal TLCDs for reducing fatigue and extreme loads of a floating offshore wind turbine. Eng. Struct. **209**, 110260 (2020)
21. Wang, Q., Qiao, H., De Domenico, D., Zhu, Z., Tang, Y.: Seismic performance of optimal Multi-Tuned Liquid Column Damper-Inerter (MTLCDI) applied to adjacent high-rise buildings. Soil Dyn. Earthq. Eng. **143**, 106653 (2021)
22. Li, H.N., Huo, L.S.: Optimal design of liquid dampers for structural torsion coupled vibration. Chin. J. Comput. Mech. **22**(2), 129–134 (2005)
23. Amini, F., Ghaderi, P.: Hybridization of harmony search and ant colony optimization for optimal locating of structural dampers. Appl. Soft Comput. **13**(5), 2272–2280 (2013)
24. Miguel, L.F.F., Lopez, R.H., Miguel, L.F.F., Torii, A.J.: A novel approach to the optimum design of MTMDs under seismic excitations. Struct. Control. Health Monit. **23**(11), 1290–1313 (2016)
25. Nigdeli, S.M., Bekdaş, G., Yang, X.S.: Optimum tuning of mass dampers by using a hybrid method using harmony search and flower pollination algorithm. In: Harmony Search Algorithm. ICHSA 2017. Advances in Intelligent Systems and Computing, vol. 514, pp. 222–231. Springer, Singapore (2017)
26. Sharma, V., Arun, C.O., Krishna, I.P.: Development and validation of a simple two degree of freedom model for predicting maximum fundamental sloshing mode wave height in a cylindrical tank. J. Sound Vib. **461**, 114906 (2019)
27. Rao, R.V., Savsani, V.J., Vakharia, D.P.: Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput. Aided Des. **43**(3), 303–315 (2011)
28. Rao, R.V.: Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int. J. Ind. Eng. Comput. **7**(1), 19–34 (2016)
29. MathWorks Announces Release 2018a of the MATLAB and Simulink Product Families. In: MathWorks. Natick, MA, US (2018)
30. FEMA P-695: Quantification of Building Seismic Performance Factors. Washington, D.C (2009)
31. Ocak, A., Bekdaş, G., Nigdeli, S.M.: A metaheuristic-based optimum tuning approach for tuned liquid dampers for structures. Struct. Design Tall Spec. Build. **31**(3), e1907 (2022)

A Survey on Optimization of Multi-criteria RBFN

Abdul-Majeed Ayebire(✉), Shelly Arora, and Mukhdeep Singh Manshahia

Department of Mathematics, Punjabi University Patiala, Punjab, India
ayebiremaj@gmail.com, aroshelly@pbi.ac.in

Abstract. Literature review plays a very important role to the success of any research work. It involves perusing works done by other scholars in the research area. It forms the basis for understanding the research topic and helps to bring to light recent works done in the research area and areas still seeking for academic attention. Literature review surveys academic articles, books, and other sources relevant to a precise area of research. These sources provide adequate information to assist a researcher have a good overview of the research area and form an opinion as to which direction his/her research should take. These sources of literature have their different characteristics as well as their individual strengths and weaknesses. In view of this, it is recommended that, the various sources are used to complement each other in a research process. In this article, the authors examine different works done in the area of Radial Basis Functions Networks (RBFs) by reviewing different sources of literature. Radial basis function networks (RBFNs) $\varphi(r)$ are univariate continuous real valued functions whose output depends on only the distance between a point and a fixed point called centre. They are regarded as a class of feed forward networks with universal approximation capabilities. This makes RBFNs an exciting area in mathematics and thus beginning to receive lots of attention.

Keywords: Neural network · Radial basis function networks · Multi-criterions optimization · Learning · Classification · Clustering · Approximation

1 Introduction

Literature is an assembly of academic writings on a subject. Literature therefore involves the process of scanning, identifying and bringing relevant scholarly writings on the topic of interest into one document. The act of undertaking this academic exercise is what is referred to as literature review. [1] defined literature review as the collection of accessible documents (both published and unpublished) on the topic, which contain information, ideas, data and evidence written from a particular perspective to justify certain aims or express certain views on the nature of the topic and how it is to be examined, and the effective evaluation of these documents in relation to the proposed research.

Literature review forms an important part of an academic document where previous works done by other writers in the research area are highlighted. That is thoroughly reading, critically analysing and briefly presenting academic materials on a given topic.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 205–213, 2024.
https://doi.org/10.1007/978-3-031-50151-7_20

It is important to state that, review of literature is not entirely a case of reading through other scholars' works and simply summarising what is in their work. As espoused by [2], literature reviews should be concise and give a clear picture of the state of knowledge and major questions in the research area. It involves reading and evaluating previous works and establishing the links with your research. According to [3], literature review should be a coherent argument that leads to the description of a proposed study.

Review of literature is critical to any academic research as it forms the basis upon which the researcher abreast him/herself with the subject area by engaging previous authors through their write ups. The researcher through the review of literature stands a better chance of appreciating the level of work done in his/her research area, the deficiencies or knowledge gaps that need to be attended to. As [4] aptly put it, the purpose of literature review is to locate the research project, to form its context or background, and to provide insights into previous work. Certainly literature view is the best starting point of generating research ideas, questions and hypothesis and therefore it is considered the first part of every scientific research.

Literature is regarded a continuous activity which begins with your encounter with the first scholarly material related to your research and ends the day you complete your work. Literature review may vary in scope and nature depending on the field and nature of research being carried out. Review of literature maybe stand-alone review that is, it involves providing a summary or overview of a research area. Systematic literature review on the other hand uses analytical methods to collect and analyse secondary data for high quality evidence. Literature review may also be dedicated recursive. That is a complete chapter may be dedicated to the literature review or it may be interspersed throughout the written document. Clearly the style of literature approach a researcher chooses to adopt will depend on the purpose of his/her research, the field of research, and the requirements of his/her research audience among a number of other compelling factors.

1.1 Sources of Literature Review

To write a successful literature review, it is important to undertake a literature search. According to [5], literature search is a detailed and organize search of all kinds of published literature in order to identify as much information as possible that are relevant to a particular topic. One way to start preparing for a literature review is to consider where your proposed review fits into Cooper's Taxonomy of Literature Review. That is focus, goal, perspective, coverage, organization and audience of your proposed research [6]. A good literature search will help reduce the time used in looking for information. It will also help clarify the scope of the research and also in finding a gap or niche [7] in the literature that your research can fill. It also plays an important role in identifying important and influential published works in the research area and also identifying relevant publishers, journals and conferences. Sources of literature can be classified into three. These are primary, secondary and tertiary sources of literature. These sources may be published or unpublished.

1.1.1 Primary Source

This refers to original or first-hand studies based on direct observation, experimental methods, interviews or use of statistical records. In other words, primary sources of literature refer to research materials by the original author. Primary source of literature includes reports, theses, emails, and conference proceedings. They require a high level of detail with little time required to publish.

1.1.2 Secondary Source

This source of literature reviews refers to literature derived from the interpretations and evaluations of literature from primary sources. That is the summary authored by a researcher other than the original researcher. This source of literature also includes journals, newspapers, reviewed articles, books and some government publications. The level of detail and time needed to publish are considered medium.

1.1.3 Tertiary Source

This source of literature is a combination of primary and secondary sources of literature. They include indexes, databases, catalogues, encyclopaedias, dictionaries, bibliographies, citation indexes. The detail requirement is low and the time needed to publish then is quite lengthy (Fig. 1).

2 Radial Basis Function Neural Networks (RBFNNs)

In the field of mathematical modelling, a radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction, classification, and system control. Powell [8] used Radial Basis Function Neural Network to solve interpolation problem in a multi-dimensional environment were as many centres as data points were needed. However, [9] devised an approach that involved the use of few centres than data points. This revolutionised the practical application of RBFNs involving large samples. RBFNs has a learning algorithm that is fast as compared to other to neural networks and are useful in representing complex non-linear mapping [10]. However, [11] focused on enhancing the generalization properties of RBFNs (Fig. 2).

2.1 Structure of RBFNs

Radial basis function (RBF) networks have a basically different construction from most neural network architectures. Most neural network architecture comprises of many layers and introduces nonlinearity by frequently applying nonlinear simulation functions. RBF network on the other hand only contains an input layer, a single hidden layer, and an output layer. The input layer consists of one or more predictor variables each linked with one independent neuron. Values from the input layer re fed to neurons in the hidden

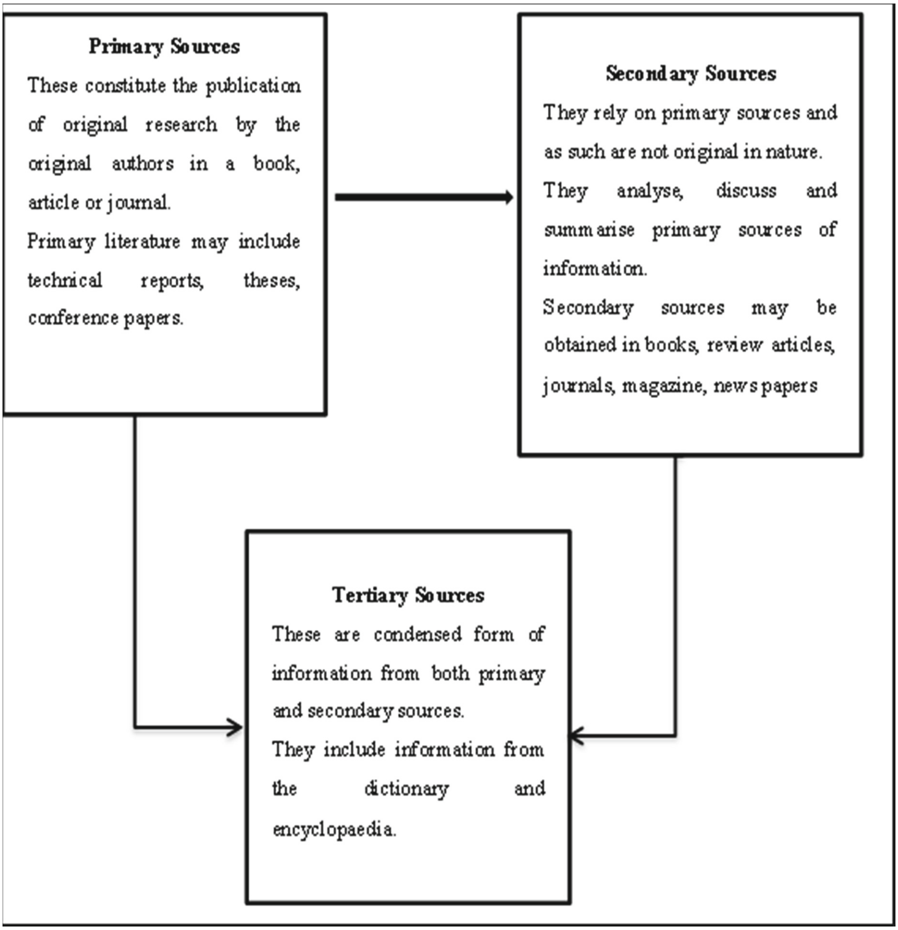


Fig. 1. Sources of literature review

layer. Neurons in the hidden layer consist of RBFs centred at a point depending on the size of the input and output variables. Hidden unit activations are given by some basis functions $\varphi(r)$. Values coming out of a neuron in the hidden layer are multiplied by some weights ($\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_n$) connected with the neurons and forwarded for summation. Here the weighted values are summed and passed out as the output of the network. Some commonly used basis functions are (Fig. 3).

Gaussian Function

$$(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad \sigma > 0 \tag{1}$$

Multi-Quadric Function

$$\varphi(r) = \sqrt{r^2 + \sigma^2}, \quad \sigma > 0 \tag{2}$$

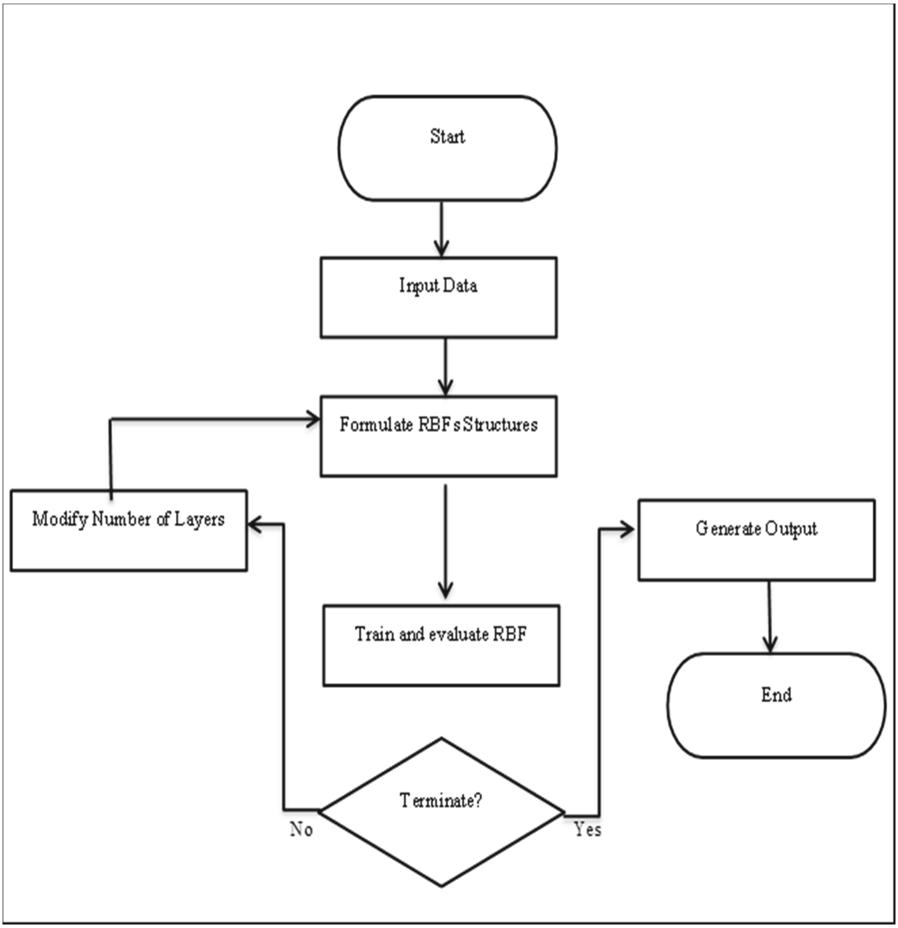


Fig. 2. Flow chart of a radial basis functions neural network (RBFNN)

Inverse Multi-Quadric

$$\varphi(r) = \frac{1}{\sqrt{r^2 + \sigma^2}}, \quad \sigma > 0 \tag{3}$$

where

$$r = \vec{x} - \overrightarrow{\mu_i} \tag{4}$$

3 Optimization of Multi-Criterions RBF Networks

Optimization of RBFNs model is the focus of current research. In particular optimization of RBFNs focuses on two objectives that is improving accuracy by minimizing error and reducing computational cost by reducing the number of RBFs as low as possible [12].

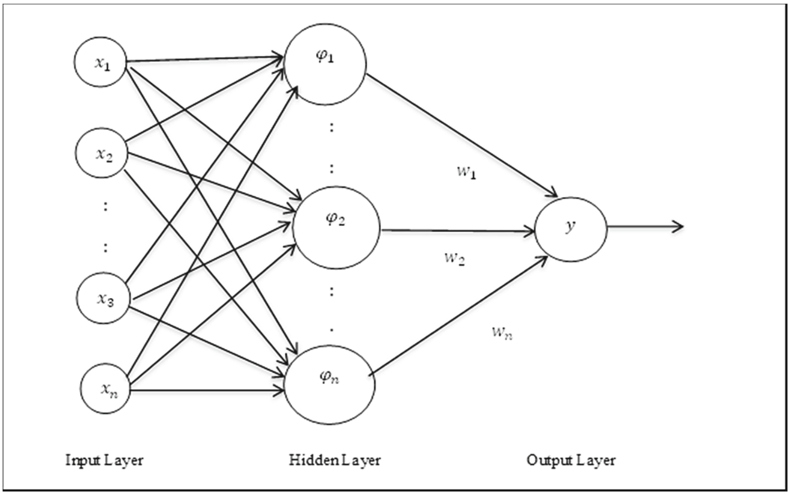


Fig. 3. The structure of a radial basis function neural network (RBFNN)

Concurrently optimizing these two objectives is known as multi-criteria optimization RBFNs.

Multi-objective optimization presents real practical challenge, however, the concept of dominance, evolutionary algorithms (EAs) [13, 14] and swarm intelligence (SI) [15] provides a means to deal with the challenges associated with multi-objective optimization problems. Optimization of size, shape and position parameters of RBFNs have been done using evolutionary algorithm [12, 16]. Multi-objective genetic algorithm has been used to identify RBFNs coupled models of humidity and temperature in a greenhouse [17]. Also [18] submitted a multi-objective Genetic Algorithm (GA) that designs RBFNs to approximate functions by minimizing network with the smallest number of neurons along with the smallest error simultaneously. A model of hierarchical rank density genetic algorithm to evolve topology and parameters of RBFNs was developed and used to optimize two criteria like prediction error and RBFs in the hidden layer for effectively predicting chaotic time series [19].

4 Applications of RBF Neural Networks

RBF networks are being used for function approximation, pattern recognition, and time series prediction problems. Such networks have the universal approximation property, arise naturally as regularized solutions of ill-posed problems and are dealt well in the theory of interpolation.

RBFNs have been applied in various areas of prediction economic factors like GDP, inflation, income and expenditures, performance prediction, for example predicting the performance and emission characteristics of a diesel engine fuelled with waste cooking oil (WCO) [20], the quality of drinking water [21] predicting weather and rainfall pattern [22, 23], air quality [24] among others.

RBFNs have been used in the health sector for medical diagnosis. RBFN have been used to accurately classify cancer tumours from breast cancer data [25], classify cancer [26], for diabetes diagnosis and classification [27–29] among other diseases.

RBFNs have also been used to demonstrate its ability for image classification and pattern recognition. It has been used for facial expressions and recognition, voice recognition and classification [30–33]. A number of researches have also been done in the areas of vehicle license plate and hand written characters' recognition [34–36].

In the field of differential equations, RBFNs have also proved to be a powerful tool. By using RBF an approximate particular solution can be achieved. [181] used modified RBFs to solve Partial Differential Equations (PDEs). Other works on the application of RBFs to Differential Equations include [37–40].

5 Conclusion

A Radial Basis Function is a real-valued function, whose value depends only on the distance from the origin. Although there are various types of radial basis functions, the Gaussian function is the most common in use. Recent studies on RBFNs are concerned with optimizing two objectives that is minimizing the model's prediction error and also keeping the number of RBFNs models as low as possible. The concept of simultaneously pursuing these two objectives is termed as the multi-criteria optimization problem in RBFNs. In multi-objective optimization, the concept of domination is very vital. By this concept, two solutions are compared on the basis of whether one dominates the solution or not.

Radial Basis Function Neural Networks have a wide scope of applications. RBFNs are useful in the field of medical diagnosis and classification of diseases, predicting economic factors, predicting weather conditions, and predicting performance. RBFNs is also applied in image, pattern, facial and character recognition. This is a useful tool for detecting fraud and fighting crime. RBFNs are applied in solving differential equation. RBFNs tend to provide superior results as compared to other methods as such it is not surprising they are receiving a lot of attention in recent times.

In the wake of the efficient results of RBFNNs in relation to other methods, the critical question then would be where lies the future of statistical tools used for predictions, classifications and forecasting. It will be interesting to see more new areas where RBFNs can be applied and how their results will compare with other methods.

Acknowledgement. Authors are grateful to Punjabi University, Patiala for providing adequate library and internet facilities.

References

1. Hart, C.: Doing A Literature Review: Releasing the Social Science Research Imagination. Sage, London (1998)
2. Bell, J.: Doing Your Research Project: A Guide for First-Time Researchers in Education, Health and Social Science, 5th edn. Open University Press, Maidenhead (2010)

3. Rudestam, K.E., Newton, R.R.: Surviving Your Dissertation: A Comprehensive Guide to Content And Process, 3rd edn. Sage Publications Inc, Thousand Oaks, CA (2007)
4. Blaxter, L., Hughes, C., Tight, M.: How to Research, 4th edn. Open University Press, Buckingham (2010)
5. Gash, S.: Effective Literature Searching for Research, 2nd edn. Gower Publishing Ltd, Aldershot (2000)
6. Cooper, H.M.: Organizing knowledge synthesis: a taxonomy of literature reviews. Knowl. Soc. **1**, 104–126 (1988)
7. Swales, J.M.: Genre Analysis: English in Academic and Research Settings. Cambridge University Press, Cambridge, UK (1990)
8. Powell, M.J.D.: Radial basis functions for multivariate interpolation: a review. In: Mason, J.C., Cox, M.G. (eds.) Algorithms for the Approximation, pp. 143–167. Clarendon Press (1987)
9. Broomhead, D.S., Lowe, D.: Multivariable functional interpolation and adaptive networks. Complex Syst. **2**, 321–355 (1988)
10. Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: A Distribution–Free Theory of Nonparametric Regression. Springer Verlag, New York, NY (2002)
11. Bishop, C.: Improving the generalization properties of radial basis function neural networks. Neural Comput. **3**(4), 579–588 (1991)
12. González, J., Rojas, I., Pomares, H., Ortega, J.: RBF neural networks, multiobjective optimization and time series forecasting. In: Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence. IWANN 2001. Lecture Notes in Computer Science, vol. 2084, pp. 498–505. Springer, Berlin, Heidelberg (2001)
13. Deb, K.: Evolutionary multi-criteria optimization. In: Proceedings of GECCO (Companion), pp. 1155–1162 (2010)
14. Coello, C.A.C.: Evolutionary multiobjective optimization. WIREs Data Min. Knowl. Discovery **1**(5), 444–447 (2011)
15. Coello, C.A.C., Dehuri, S., Ghosh, S. (eds.): Swarm Intelligence for Multi-Objective Problems in Data Mining. Springer-Verlag, Berlin, Heldelberg (2009)
16. Gonzàlez, J., et al.: Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation. IEEE Trans. Neural Networks **14**(6), 1478–1495 (2003)
17. Ferreira, P.M., Ruano, A.E., Fonseca, C.M.: Evolutionary multiobjective design of radial basis function networks for greenhouse environmental control. In: IFAC Proc. **38**(1), 63–68 (2005)
18. Guillén, A., Rojas, I., González, J., Pomares, H., Herrera, L.J., Paechter, B.: Improving the performance of multi-objective genetic algorithm for function approximation through parallel islands specialisation. In: AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science, vol. 4304, pp. 1127–1132. Springer, Berlin, Heidelberg (2006)
19. Yen, G.G.: Multi-objective evolutionary algorithm for radial basis function neural network design. In: Multi-Objective Machine Learning. Studies in Computational Intelligence, vol. 16, pp. 221–239. Springer, Berlin, Heidelberg (2006)
20. Shiva, K., Pai, P.S., Rao, B.R.S.: Radial-basis-function-network-based prediction of performance and emission characteristics in bio diesel engine run on WCO ester. Adv. Artif. Intell. **2012**(610487), 1–7. Hindawi Publishing Corporation (2012)
21. Bouamar, M., Ladjal, M.: A comparative study of RBF neural network and SVM classification techniques performed on real data for drinking water quality. In: 2008 5th International Multi-Conference on Systems, Signals and Devices, pp. 1–5. IEEE (2008)
22. Santhanm, T., Subhajini, A.C.: An efficient weather forecasting system using radial basis function neural network. J. Comput. Sci. **7**(7), 962–966 (2011)

23. Wu, J.: Prediction of rainfall time series using modular RBF neural network model coupled with SSA and PLS. In: Intelligent Information and Database Systems. ACIIDS 2012. Lecture Notes in Computer Science, vol. 7197, pp. 509–518. Springer, Berlin, Heidelberg (2012)
24. Bhattacharya, S., Shahnawaz Sk: Using Machine Learning to Predict Air Quality Index in New Delhi (2021). arXiv:2112.05753, arXiv:2112.05753v1, https://doi.org/10.48550/arXiv.2112.05753
25. Subashini, T.S., Ramalingam, V., Palanivel, S.: Breast mass classification based on cytological patterns using RBFNN and SVM. Expert Syst. Appl. **36**(3, Part 1), 5284–5290 (2009)
26. Chu, F., Wang, L.: Applying RBF neural networks to cancer classification based on gene expressions. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 1930–1934. IEEE (2006)
27. Kayaer, K., Yildirim, T.: Medical diagnosis on Pima Indian diabetes using general regression neural networks. In: Proceedings of the International Conference on Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP), vol. 181, p. 184 (2003)
28. Temurtas, H., Yumusak, N., Temurtas, F.: A comparative study on diabetes disease diagnosis using neural networks. Expert Syst. Appl. **36**(4), 8610–8615 (2009)
29. Venkatesan, P., Anitha, S.: Application of a radial basis function neural network for diagnosis of diabetes mellitus. Curr. Sci. **91**(9), 1195–1199 (2006)
30. Thakur, S., Sing, J.K., Basu, D.K., Nasipuri, M., Kundu, M.: Face recognition using principal component analysis and RBF neural networks. In: 2008 First International Conference on Emerging Trends in Engineering and Technology, pp. 695–700. IEEE (2008)
31. Dhanalakshmi, P., Palanivel, S., Ramalingam, V.: Classification of audio signals using SVM and RBFNN. Expert Syst. Appl. **36**(3, Part 2), 6069–6075 (2009)
32. Rosenblum, M., Yacoob, Y., Davis, L.S.: Human emotion recognition from motion using a radial basis function network architecture. IEEE Trans. Neural Networks **7**(5), 1121–1138 (1996)
33. Arad, N., Dyn, N., Reisfeld, D., Yeshurun, Y.: Image warping by radial basis functions: applications to facial expressions. CVGIP Graph. Models Image Process. **56**(2), 161–172 (1994)
34. Shan, B.: License plate character segmentation and recognition based on RBF neural network. In: 2010 Second International Workshop on Education Technology and Computer Science, pp. 86–89. IEEE (2010)
35. Sitamahalakshmi, T., Babu, V., Medapati, J., Chandra, M., Mouli, C., Mouli, K.V.V.C.: Performance comparison of radial basis function networks and probabilistic neural networks for Telugu character recognition. Glob. J. Comp. Sci. Technol. **11**(4), 9–15 (2011)
36. Lee, Y.: Handwritten digit recognition using k-nearest neighbor, radial basis function and backpropagation neural networks. Neural Comput. **3**(3), 440–449 (1991)
37. Franke, C., Schabck, R.: Convergence orders estimates of meshless collocation methods using radial basis functions. Adv. Comput. Math. **8**, 381–399 (1998)
38. Schaback, R., Wendland, H.: Using compactly supported radial basis functions to solve partial differential equations. Trans. Model. Simul. **22**, 311–324 (1999)
39. Kansa, E.J.: Multiquadrics–a scattered data approximation scheme with applications to computational fluid dynamics-I surface approximations and partial derivative estimates. Comput. Math. Appl. **19**(8–9), 127–145 (1990)
40. Kansa, E.J.: Multiquadrics–a scattered data approximation scheme with application to computational fluid dynamics-II solutions to parabolic, hyperbolic and elliptic partial differential equations. Comput. Math. Appl. **19**(8–9), 147–161 (1990)
41. Ling, L., Kansa. E.J.: A least-squares pre-conditioner for radial basis functions collocation methods. Adv. Comput. Math. **23**, 31–54 (2005)

Load Balancing in Cloud Environment Using Different Optimization Algorithms and Open-Source Platforms: A Deep Picture

Akashbhai Dave[1] and Hetal Chudasama[2,3](✉)

[1] Ph.D. Scholar, The Charutar Vidya Mandal (CVM) University, Vallabh Vidhyanagar, Anand, Gujarat, India
akashbhai.dave@cvmu.edu.in
[2] Department of Information Technology, G H Patel College of Engineering and Technology, Vallabh Vidhyanagar, Anand, Gujarat, India
hetal.chudasama@cvmu.edu.in
[3] The Charutar Vidya Mandal (CVM) University, Vallabh Vidhyanagar, Anand, Gujarat, India

Abstract. Cloud computing is a very vast area and become very popular in IT industries and academics. It has many challenges that load balancing is a critical/Important issue. Load balancing is a very interesting topic for researchers. In this article, we are studying cloud computing, load balancing, Virtualization, Optimization algorithms of evolutionary and swarm-based algorithms, and open-source Cloud Platforms, with different-different parameters.

Keywords: Load balancing · Cloud computing · Open-source cloud platforms · Optimization algorithm

1 Introduction

Cloud computing is expanding quickly as an effective worldview introducing the on-request framework, stage, and programming administrations to customers [1, 2]. Load Balancing is disseminating the powerful responsibility similarly among every one of the hubs to keep away from the status that a few hubs are over—Balancing while others are under-stacked [1–5]. Numerous calculations have been proposed to play out this assignment [4, 6].

2 Cloud Computing

Cloud computing has arisen as an innovation that lubes undertakings by the unique allotment of VM [4]. A cloud supplier should confront numerous challenges [7]. Cloud computing addresses various difficulties at an expanding number of clients because the

The authors have no conflicts of interest to declare that are relevant to the content of this article. No funding was received to assist with the preparation of this manuscript.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 214–222, 2024.
https://doi.org/10.1007/978-3-031-50151-7_21

interest in asset sharing and utilization are expanded quickly [1, 8, 9] cloud computing has become a most important part of the market because numerous associations and industries don't have the fundamental assets because of the significant expense of programming and equipment [2, 5, 8, 10]. To conquer this issue, cloud computing gives the necessary programming and equipment at a reasonable rental expense [1].

3 Virtualization

Virtualization innovation is to set up a product layer called a hypervisor on an actual equipment stage. A hypervisor is a working framework that ensures the assets of a VM Host and deals with the solicitations and reactions from VMs [4, 10, 11].

Contingent upon the application conditions, type-1 hypervisors straightforwardly run on VM Host's equipment, and type-2 run on the VM Host's OS. For additional insights regarding hypervisors [2], if it's not too much trouble, allude to them. A client program running on a VM can't straightforwardly get to the actual equipment. All things being equal, all assets are wrapped by the Virtual Machine Host's hypervisor. For the most part, different VMs are executed on and overseen by a Virtual Machine Host [11].

4 Load Balancing

Load adjusting is successfully utilized for getting a high and adequate nature of administration. Load balancer conveys customer demands across numerous workers. They utilize the idea of virtualization with the assistance of hypervisors on a VM director [7]. There are three categories [2, 12]: (1) sender-initiated, (2) receiver-initiated, and (3) symmetric, a combination of both modes is available for load balancing algorithms.

Static and dynamic are two types of LB algorithms available. In static, the adjusting component is done before execution [2]. In Dynamic, the assignments are executed progressively between all assets, and it is important to screen the current heap of the framework [1].

5 Optimization

The most adopted or proposed algorithms in cloud optimization are ACO, PSO, GA, and others. ACO gives a promising variable of errand planning and an asset to the board like security versatility, time, execution, and cost advancement, beating other bio-roused calculations [6]. Numerous bio-inspired algo. Demonstrate their effectiveness in load balancing frameworks like ant colonies and honeybees [1]. Bio-inspired algorithms are separated into swarm intelligence algorithms and Evolutionary algorithms [4, 11, 12].

Evolutionary algorithms are created by copying the natural behavior for determination and upgrade. These algorithms are partitioned into two subclasses: genetic programming and genetic algorithms [4, 12, 13]. Relying upon the conduct of some natural living animals, like insects, honeybees, fishes, and birds which have their approaches to find the inquiry space of the problems are called Swarm intelligence algorithms [1, 4].

5.1 Honeybee Optimization [14]

A Honeybee Behavior-based Load Balancing plot is adopted for achieving the reasonable offset of VM with a target of augmenting throughput in the distributed computing situation. It manages to adjust the assignment needs over the VM to such an extent that the holding up time is limited concerning the number of errands in the line. It is proposed for forestalling under and over stacking of VM's to ensure critical reaction time and ideal machine usage. It is recognized to diminish the pausing and mean execution seasons of undertakings accessible in the line.

5.2 Particle Swarm Optimization [3, 4, 7, 9, 11, 12, 15–17]

This calculation is a bio-motivated calculation that is reached out from the social conduct of birds rushing to look for food [3, 5, 8]. In this, every particle in the group of swarm goes about as an answer with four vectors, its speed, the best situation found by other best arrangements among all particles in the populace, the best position tracked down, and its current position.

It's an exceptionally modern bionic heuristic, keen optimization algo. That duplicate the swarm-based conduct of creatures. Its calculation isn't productive in addressing differential limitation issues. With the benefits of concurrent allotment, extensibility, ease of perception, amazing flexibility, with great attributes in unique conditions, PSO effectively beats various issues identified with blending optimization [3].

5.3 Ant Colony Optimization [6, 8, 14, 15, 17, 18]

Ant Colony Optimization (ACO) is an irregular choosing calculation that relies upon the ant's conduct. It relies upon looking generally for food by its smell trails for association and back to their home through the briefest course by the centralization of smell. The grouping of the smell begins vanishing. The force of the smell relies upon the integrity and distance of the food. The ACO adjusts to dynamic conditions. It is incredibly acceptable in adaptation to non-critical failure and adaptability which works on the exhibition of the framework.

5.4 Genetic Algorithm [10, 15, 17, 19]

It's an algo. That attempts to adjust the weight on cloud assets while attempting to diminish the end time for the assignment set being referred to. It is a continuous inquiry algo. Depending on the cycles of normal choice and genetics. A straightforward GA comprising of triple cycles: (1) accessibility, (2) genetic, and (3)substitution.

5.5 Osmos Optimization [3, 14]

Osmosis LB Model reassigns errands to a progression of VM, fully intent on changing the heap. The arrangement of LB is scattered and is upheld by the insect-like specialists that information lopes lead on a Chord overlay. Every server farm communicates with a rundown and can execute one or more undertakings at the same time with various attributes of the implementation [3].

5.6 Bee Colony Optimization [3]

An enhancement strategy propelled by the decision-making measure for LB utilizing a counterfeit honeybee settlement algorithm. Honeybee settlement streamlining brings the choosing cycle by searching for the best food alternatives across various freedoms. The decision-making measure is reliant upon the Swarm [3].

6 Literature Survey

In [1], in this paper, hybrid metaheuristics propose a procedure that joins osmotic behavior with bio–inspired LB algorithms. It empowers the programmed arrangement of the VM that is relocated through cloud foundations. It conquers the disadvantages of the current bio-inspired algorithms in accomplishing load accommodation between actual machines.

- Issues Solved: Reduce energy consumption, VMM counts, Reduce the number of shutdown costs, enhance the Quality of Services (QoS)
- Platform Used: CloudSim API 3.0.3 (Simulation)
- Comparison Done with: ACO, ABC, H_BAC, IQR, Median Absolute Deviation, and Local Regression.
- Result: Improves the number of VM migrations and energy consumption, compared to other algorithms reduces the number of host's shutdowns.

[3] This survey paper's goal is to introduce a basic investigation of already existing strategies of LB, to talk about different LB boundaries i.e., overhead, response time, scalability, resource usage, throughput, performance, migration time, fault tolerance, and so forth the exploration paper additionally talks about the issues of LB and distinguishes the requirement for an original LB algorithm that utilizes FT measurements.

[6] This examination proposes Spanning Tree with Ant Colony Optimization using Finite Automata, a machine-learning idea for overseeing cloud assets in a multi-level climate. The general target is to reduce energy consumption through information position influence and virtual machine combination. The proposed method's productivity was benchmarked on four execution measurements.

- Issues Solved: Reduce energy consumption.
- Platform Used: Simulation
- Comparison Done with: ICGA, RACO, ICSA, and FAACOST
- Result: Setting Benchmark.

[4] The proposed strategy for the current investigation has been focused on the opportune situation of VM on the physical machine dependent on planning undertakings to boost benefits just as diminishing working expenses and forestalling execution misfortune in assistance conveyance. The strategy proposes a strategy utilizing a meta heuristic approach just as multi objective PSO & PSO algorithms to expand benefit and execution, decline working expenses, and streamline the usage of assets in cloud computing.

- Issues Solved: Improving task scheduling.
- Platform Used: MATLAB R2015a (Simulation)

- Comparison Done with: PSO.
- Result: Task Scheduling Archived.

[12] In this paper, the proposed procedure is fitting for handling the asset optimization challenges, because of the capacity to change over the load balancing issue into an optimization issue (decreasing lopsidedness cost). In this examination, joining the consequences of the particle swarm genetic optimization algorithm and utilizing a mix of benefits of these two algorithms lead to the enhancement of the issues and presenting a reasonable answer for load balancing activity, because in the proposed approach, rather than aimlessly relegating the underpinning crowd in the inheritable algorithm, the stylish outgrowth is secured by putting the underpinning crowd. * eLBPSGORA fashion is varied with PSO, GA, and mongrel GA–PSO. Prosecution cost, cargo balancing, and makespan have been assessed and our fashion has performed more compared to relative strategies.

- Issues Solved: Improve Resource Allocation, Energy Consumption
- Platform Used: Simulation
- Comparison Done with: Hybrid GA-PSO and PSO, GA.
- Result: Proposed method is more efficient in execution cost and make span.

[7] The proposed strategy centers around the mutation-based Particle Swarm algorithm to balance load among the server farms. Here an effective load balancing algorithm is created to limit execution boundaries like Make Span time and further develop the wellness work in distributed computing. The proposed technique has a reason to emphasize the uncovering ability of the inquiry cycle. By and large, very good quality workers contain an immense measure of registering power and storing assets. These workers typically speak with each other with the assistance of a high transmission capacity intercommunication organization. In this manner, transmission delay doesn't assume a lot of critical part in distributed computing. In this climate, every client uses cloudlets with the assistance of the Internet.

- Issues Solved: Make Span
- Platform Used: CloudSim 3.0.3 (Simulation)
- Comparison Done with: PSO.
- Result: Proposed technique gives an optimized Make Span rather than the existing PSO.

[11] In this paper, they take on a systematic literature review technique to choose contemplates dependent on outlined Research Questions and predefined incorporation/avoidance standards from online electronic information bases. 39 articles were chosen out of 364 articles. Those articles that answer their outlined examination questions are chosen for investigation. The examination thinks about different benefits, disadvantages, properties, reproduction and assessment instruments, work targets of PSO, and gets ready characterization dependent on PSO altered and PSO crossover. At last, a future exploration heading is introduced.

[19] In This Paper, they propose another strategy for work process planning in the cell registering environmental elements which utilize the optimization–based thoroughly booking with climate well–disposed result. The compact system is impressively significantly less steady than the wired condition; energy should be respected at the arranging

time. Occupation planning for cell frameworks hence requires a solid machine life-sized model that can contain this load of components. To collect the lively and portable nature of assets, the accessibility should be assessed in the event of booking. In Grid Booster Algorithm's is utilized to schedule the way to deal with permit brilliant load balancing.

- Issues Solved: Resources Utilization
- Platform Used: Simulation
- Comparison Done with: GA.
- Result: Proposed planning approach will join faster in large tasks and increase each perfect association.

[10] In this paper, the authors, present a two-stage genetic mechanism for the migration-based load balance of VMH in cloud computing. Past methods generally accept this issue as a task optimization issue and just consider the current VM Host's loads; in any case, disregarding loads of VMH after adjusting, these methods can just acquire restricted viability in genuine conditions.

- Issues Solved: Performance
- Platform Used: Jnet
- Comparison Done with: Gene expression programming (GEP)
- Result: Their method outperforms previous methods.

[5] In this examination, they have depicted different load balancing methods in cloud computing conditions. A framework engineering, with unmistakable models for the host, Virtual Machine is portrayed. They have clarified different execution boundaries recorded in the above tables that assess the framework execution. The estimation of the make span and energy utilization of the framework is clarified exhaustively.

[15] This paper proposes an original technique of dynamic balancing of load among the virtual machines using hybridization of modified Particle swarm optimization and a bettered Q–learning algorithm. A hybridization cycle is done to change the speed of the MPSO with the gbest and pbest dependent on the best activity created through the bettered Q–literacy. The point of hybridization is to improve the exhibition of the virtual machine by balancing the load among the Virtual Machines, boost the throughput of Virtual Machines, and keep up with the harmony between requirements of assignments by streamlining the holding up season of undertakings. The strength of the algorithm has been approved by differing the effects of the QMPSO obtained from the reduplication commerce with the current load balancing and scheduling algorithm. The correlation of the reduplication and genuine stage result shows our proposed algorithm is beating its rival.

- Issues Solved: Make span, throughput, energy utilization, reduces the waiting time
- Platform Used: Cloud Sim (Simulation)
- Comparison Done with: IPSO
- Result: Reduces the waiting time of the tasks, improves the make span, throughput, energy utilization.

[8] The author presents a Multi-resource Load Balancing Algorithm for cloud environment and algorithm depends on Ant Colony Optimization. The proposed technique targets make span cost.

- Issues Solved: Make Span
- Platform Used: Simulation
- Comparison Done with: ACO, GA ACO, Hybrid
- Result: Setting Benchmark.

[18] In this paper, they approved their proposed approach by directing broad reproductions on various traffic examples and situations with various thresholds. The outcomes show that the proposed routing technique extensively limits the network energy utilization, particularly for clogged deals with mice-type streams. It can give compelling connection load adjusting while at the same time fulfilling the clients' QoS prerequisites.

- Issues Solved: Energy consumption
- Platform Used: Simulation
- Comparison Done with: Default controller algorithm
- Result: Improve performance.

[13] The resources used state is considered in each creation for gatherings. In the wake of finding a defective solution for each gathering, the proposed technique adds every one of the tricky responses for clusters into the last task map. Finally, MOSPSO endeavors to change the heaps during the last task map.

- Issues Solved: Make Span
- Platform Used: CloudSim
- Comparison Done with: PSO, IPSO
- Result: Adaptive heuristic-based models give a Very good outcome.

[14] The authors proposed algo. includes the worldwide investigation ability of ABC and the nearby double–dealing capability of MBOA for the successful distribution of client errands to VMs. It centers around network and computing resources to forestall discontinuity and superfluous increase in the assignment finishing times as both ought to be possibly investigated for a superior resource distribution measure.

- Issues Solved: Make Span
- Platform Used: Cloud Sim (Simulation)
- Comparison Done with: DLB Throttled-LBA WRR
- Result: Reduce Migrations.

[20] The author shows contrasts between OpenStack and OpenNebula from Different perspectives. They likewise present recommendations for execution. They accept that this assessment will assist individuals with understanding their capacities, objectives, and varieties. The proposals will bring about better utilization of OpenStack and OpenNebula.

[9] This strategy sends the Firefly (FF) algorithm to limit the pursuit space while the IMPSO procedure is carried out to recognize the improved reaction. The IMPSO algorithm works by choosing the worldwide best (gbest) molecule with a little distance of a highlight line. With the utilization of least separation from a highlight line, the gbest molecule applicants could be chosen. The proposed FIMPSO algorithm accomplished a compelling normal burden for making and improving the fundamental estimates like appropriate asset use and reaction season of the errands.

- Issues Solved: Memory utilization
- Platform Used: Cloud Sim
- Comparison Done with: RR, FCFS, SJF, GA, IPSO, Firefly, FFIPSO
- Result: Average response time, Maximize CPU utilization, Maximize memory utilization, reliability, and throughput.

[16] The authors proposed modified PSO task scheduling to plan tasks over the accessible cloud resources that minimize the make span and boost resource utilization. Their proposed scheduling algorithm is carried out by utilizing the CloudSim simulator.

- Issues Solved: Make span, resource utilization.
- Platform Used: XEN
- Comparison Done with: PSOBTS, L-PSO, and DLBA
- Result: Reducing make span.

[17] In This article, they examine four distinct nature-inspired load balancing algorithms for the cloud environments and look at their performances as far as normal TRT and normal DCPT. The outcomes are looked at for changing quantities of DC and UBS.

7 Conclusion

According to, the user requests and cloud Architect, the cloud computing system is doled out with some load. Allocation or scheduling of user requests is an NP-hard optimization issue. Circumstances like overloaded and underloaded cause diverse system failures concerning resource consumption, machine failure, execution time, and so forth, load balancing is needed to beat all referenced issues. Load balancing is the component of identifying underloaded and overloaded hubs and afterward balancing the load among them. In this paper, a concise clarification of considered Optimization in the writing and its belongings is introduced.

References

1. Gamal, M., Rizk, R., Mahdi, H., Elnaghi, B.E.: Osmotic bio-inspired load balancing algorithm in cloud computing. IEEE Access **7**, 42735–42744 (2019)
2. Bist, M., Wariya, M., Agarwal, A.: Comparing delta, open stack and Xen Cloud Platforms: a survey on open source IaaS. In: 2013 3rd IEEE International Advance Computing Conference (IACC), pp. 96–100. IEEE (2013)
3. Shahid, M.A., Islam, N., Alam, M.M., Su'ud, M.M., Musa, S.: A comprehensive study of load balancing approaches in the cloud computing environment and a novel fault tolerance approach. IEEE Access **8**, 130500–130526 (2020)
4. Yousefipour, A., Rahmani, A.M., Jahanshahi, M.: Improving the load balancing and dynamic placement of virtual machines in cloud computing using particle swarm optimization algorithm. Int. J. Eng. **34**(6), 1419–1429 (2021)
5. Mishra, S.K., Sahoo, B., Parida, P.P.: Load balancing in cloud computing: a big picture. J. King Saud Univ. Comput. Inf. Sci. **32**(2), 149–158 (2020)
6. Aliyu, M., et al.: Management of cloud resources and social change in a multi-tier environment: a novel finite automata using ant colony optimization with spanning tree. Technol. Forecast. Soc. Change **166**, 120591 (2021)

7. Agarwal, R., Baghel, N., Khan, M.A.: Load balancing in cloud computing using mutation based particle swarm optimization. In: 2020 International Conference on Contemporary Computing and Applications (IC3A), pp. 191–195. IEEE (2020)
8. Muteeh, A., Sardaraz, M., Tahir, M.: MrLBA: multi-resource load balancing algorithm for cloud computing using ant colony optimization. Cluster Comput. **24**, 3135–3145 (2021)
9. Devaraj, A.F.S., Elhoseny, M., Dhanasekaran, S., Lydia, E.L., Shankar, K.: Hybridization of firefly and improved multi-objective particle swarm optimization algorithm for energy efficient load balancing in cloud computing environments. J. Parallel Distrib. Comput. **142**, 36–45 (2020)
10. Hung, L.-H., Wu, C.-H., Tsai, C.-H., Huang, H.-C.: Migration-based load balance of virtual machine servers in cloud computing by load prediction using genetic-based methods. IEEE Access **9**, 49760–49773 (2021)
11. Singh, G., Chaturvedi, A.K.: Particle swarm optimization-based approachesfor cloud-based task and workflow scheduling: a systematic literature review. In: 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), pp. 350–358. IEEE (2021)
12. Mirmohseni, S.M., Javadpour, A., Tang, C.: LBPSGORA: create load balancing with particle swarm genetic optimization algorithm to improve resource allocation and energy consumption in clouds networks. Math. Probl. Eng. **2021**(5575129), 1–15 (2021)
13. Malarvizhi, N., Aswini, J., Sasikala, S., Chakravarthy, M.H., Neeba, E.A.: Multi–parameter optimization for load balancing with effective task scheduling and resource sharing. J. Ambient Intell. Human. Comput. **13**(Suppl 1), S75 (2022)
14. Janakiraman, S., Priya, M.D.: Improved artificial bee colony using monarchy butterfly optimization algorithm for load balancing (IABC-MBOA-LB) in cloud environments. J. Netw. Syst. Manage. **29**(4), 1–38 (2021)
15. Jena, U.K., Das, P.K., Kabat, M.R.: Hybridization of meta-heuristic algorithm for load balancing in cloud computing environment. J. King Saud Univ. Comput. Inf. Sci. **34**(6, Part A), 2332–2342 (2022)
16. Pradhan, A., Bisoy, S.K.: A novel load balancing technique for cloud computing platform based on PSO. J. King Saud Univ. Comput. Inf. Sci. **34**(7), 3988–3995 (2022)
17. Arulkumar, V., Bhalaji, N.: Performance analysis of nature inspired load balancing algorithm in cloud environment. J. Ambient. Intell. Humaniz. Comput. **12**(3), 3735–3742 (2021). https://doi.org/10.1007/s12652-019-01655-x
18. Torkzadeh, S., Soltanizadeh, H., Orouji, A.A.: Energy-aware routing considering load balancing for SDN: a minimum graph-based Ant Colony Optimization. Clust. Comput. **24**(3), 2293–2312 (2021). https://doi.org/10.1007/s10586-021-03263-x
19. Santhosh, J., Aruna, R., Balamurugan, P., Arulkumaran, G.: An efficient job scheduling and load balancing methods using enhanced genetic algorithm. Eur. J. Mol. Clin. Med. **7**(8), 4502–4510 (2020)
20. Wen, X., Gu, G., Li, Q., Gao, Y., Zhang, X.: Comparison of open-source cloud management platforms: OpenStack and OpenNebula. In: 2012 9th International Conferenceon Fuzzy Systems and Knowledge Discovery, pp. 2457–2461. IEEE (2012)

In-Depth Analysis on the Realm of Cloud Computing

Jeric Dumangas(✉), Ma. Suzette Cagbabanua, Joshua Diaz, Ana Maia Ignacio, Joshua Santos, Larry Ivan Young, and Jo-Ann Magsumbol(✉)

Polytechnic University of the Philippines, Manila, Philippines
jericdumangas30jerboy@gmail.com, javmagsumbol@pup.edu.ph

Abstract. The goal is to systematically evaluate the trends, applications, factors, and impacts of the acceptance of cloud computing in today's time. The factors, applications, as well as trends in the global and local setting, were examined in this study to reveal the impacts of the said topic as the Philippines is currently transitioning to become a major data center for cloud computing in Asia. Moreover, policy recommendations derived and based on the existing policies from other countries adopting the cloud computing technology are also presented for the ongoing innovation plan in several different sectors, namely agriculture, economy, industry, labor, and society.

Keywords: Cloud computing · Data storage · Cloud service providers · Virtual machines · Web services · Utility computing · On-demand computing

1 Introduction

Cloud computing (CC) holds data and applications which run on a remote server rather than on the local computer or any other hardware. Users may then access the data and programs saved in the cloud over the internet; thus, the workload is now kept on the cloud. Cloud computing is a big building full of servers, composed of computers that provide services on behalf of customers. These buildings are gigantic with a large data center having servers extending in every direction. The servers run programs, store data, process data, and host websites, etc. [1].

Cost is one of the vital aspects that institutions examine in the adoption of cloud computing technology. A person or company can save money using cloud computing (CC) in terms of set-up cost in hardware, software, facility upkeep, and other costs associated with keeping their own data center. Utilizing the cloud as an option would be more cost-effective [2]. Dependability is another aspect in cloud computing usage since the cloud provider manages everything from backup up to disaster recovery of all data. Also, CC has multiple redundant sites that serve as a backup in case data centers fail, therefore, preventing any system downtime. Scalability is an additional feature of CC where service providers offer Pay-as-you-go services.

Both ends of the system—front and back ends—are useful sections to make when discussing cloud computing technology. The front end is the user or client section,

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 223–233, 2024.
https://doi.org/10.1007/978-3-031-50151-7_22

and the "cloud" is the back section. Types of CC include public cloud (PuC), private cloud (PrC), and hybrid cloud (HyC). In PuC, also referred as external cloud, sharing of resources takes place through internet, and is managed by providers. Private cloud (PrC) is also known as internal cloud, services are given in exclusivity to a particular company thus ensuring security, data control and service quality. Hybrid clouds (HyC) is a combination of both the public and private clouds. In terms of services, cloud models are classified as IaaS, PaaS and SaaS [3, 4] (Fig. 1).

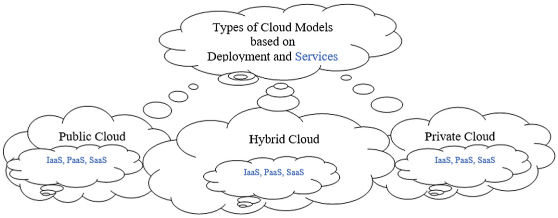


Fig. 1. Cloud computing deployment and services models

This research analysis offers brief and compact information on cloud computing: its brief history, advantages, applications, technology impacts and inhibitions, and policy recommendations.

2 Trends

In the last decade, the cloud computing landscape has shifted dramatically. Not only have there been more providers and service capabilities, but cloud infrastructure is also expanding, which was previously restricted to single-provider data centers [5].

2.1 Global Domain

Globally, cloud computing evolved in 3 phases: idea phase (also known as the client-server model), pre-cloud phase, and the cloud phase. In 1960, the client-server model was widely used until the pre-internet era, which is decentralized in nature and spread workloads or jobs across resource suppliers and service requesters [6–8]. Servers and clients are stated to interact across a computer system, and to balance demand, the server swaps resources with service users [9]. Clients cannot communicate directly with one another under this approach. Next, the Pre-cloud phase introduced the application as a service from 1999 to 2006. Finally, the transition to the cloud phase began in 2007. Several of the world's leading web firms have innovated and established forms of cloud services. By the end of this decade, it is anticipated that around 90% of the global population, or 7.5 billion people, will use and save data in the cloud. With the extensive implementation of the 5G technology, year 2022 has a projected 34.35% increase of smartphone users since 2016. As a result, cloud-based applications are anticipated to grow in popularity in the mobile application market. By 2025, it is projected that the data saved in the cloud will reach 100 zettabytes, including both in shared and personal clouds, from the 25% that was stored there in 2015. This represents 50% of the total quantity of data in existence at that time [10–12].

Cloud computing skyrocketed in 2020 and 2021 as business became virtual and companies focused on providing digital services in response to the global pandemic. It is anticipated that there will be high acceptance and growth in 2022 [13, 14].

2.2 Local Domain

As more businesses invest in data centers locally, the Philippines is one of the markets with the quickest growth. Examples of recent funding are Space DC & YCO Cloud Centers [15]. The nation's 76 million internet users, who log on for an average of over 10 h per day, provide a strong foundation for data consumption, which is anticipated to continue growing as more people utilize technology in the future, which will lead to higher data demand. Young people in the nation are expected to continue adopting e-commerce and using social media, which will support the steady rise in internet activity [16].

Digitalization drives both the government institutions and enterprises in the Philippines. CC technology is regarded as one of the drivers of digitalization, and 16% of the nation's businesses became open to employing cloud-based information and technology solutions [17]. From 2015 to 2020, it was predicted that expenditure on cloud computing will increase at a rate of 31.7% annually, hitting PhP 14.9B. Additionally, cloud computing will make up 13.3% of the market for IT services in the Philippines, more than doubling its 2015 contribution of 6.1% [18, 19].

The main drivers of the market development include COVID-19, which encourages the use of the internet, the rising use of CDNs, and gaming [19]. The market for data centers is becoming more receptive due to additional facilities supporting a wider variety of needs as compared to a few telecom operators providing services. With 12 different independent 3rd-data center facilities accounting for more than 75% of the nation's current power capacity. Manila is the Philippines' most popular site for data center operators. Like the global trend in cloud computing market size, the country will also experience an increase in data center investments from $280M in 2020, $298M in 2021, and $535M in 2026 up to $635M by 2027 [15, 20] as was depicted in Fig. 2.

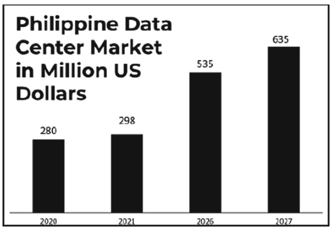


Fig. 2. The Philippine data center market in million US dollars

Cloud Readiness Index of the Philippines: Using the baseline horizon, the APAC region is estimated to provide considerable development potential. The reason that influences the cloud's usage in the APAC region includes the growing availability of trained labor and the great focus SMEs and major organizations have on entering and expanding in this region. The Philippines' position in the APAC region's cloud readiness rankings has declined two points [21]. In 2020, the Cloud Readiness Index (CRI) of the Philippines is 55.3, which dropped from ninth to eleventh place. According to criteria such as global connectivity, electricity sustainability, data center security, and cybersecurity, the CRI evaluated 14 Asia-Pacific nations. India and Indonesia's CRIs are used as comparisons while Singapore, Hong Kong, and New Zealand hold the top three spots [22, 23].

3 Applications

3.1 Online Data Storage

Depository could be used for a company database or for basic data storage, like local hard disk storage. Instead of being saved on conventional network storage, data are stored in the cloud. To supply consumers with the service, cloud storage is made up of several storage devices clustered together in a network, a shared file system, and other storing middleware [24].

3.2 Backup and Recovery

The customer or user may access resources on demand. Hardware or software can be used as a resource. Any sort of data file, app development tool, and API, among other applications, are examples. There are several options available among numerous implementations for data backup and security maintenance among numerous users. For customers to upload their sensitive and vital data, cloud computing must be dependable. When setting up any cloud, cost-effectiveness is the major priority [25].

3.3 Big Data Analysis

The phrase "big data" refers to datasets that are so huge that conventional database systems cannot store or process them. The size of the datasets is a result of the data coming from a range of new references as opposed to an old traditional structured source, such email, social media, and Internet-connected sensors. The characteristics of big data present challenges to data storage and analysis that businesses will encounter [26].

3.4 Testing and Development

CC offers product testing and development. Using Cloud Service Provider, the user can create, test, and deploy an application or product. Cloud users can use the resources, apps, and infrastructure offered by cloud providers on a pay-per-use basis. Such access may take the shape of applications or products that cloud service providers have already made available to their customers [27].

3.5 Application in Antivirus

Cloud Computing can also supply antivirus software that checks viruses and malware for the organization's system. A cloud environment is prone to malicious attacks and threats that can destroy files and applications in its environment. The traditional antivirus cannot kill new threats and other unwanted malicious files. That is why cloud computing should deploy new antivirus architecture that overcomes new virus problems [28].

3.6 Application in E-commerce

Cloud computing influences several industries, including e-learning, healthcare, and e-commerce [29]. It provides online services with strong economic value and at a reasonable price. It is without a doubt the upcoming revolution in both the corporate and Internet worlds.

The usage of cloud computing in e-commerce allows the company to virtually appear large and run widely. An e-commerce application can adapt to the shifting market conditions and demand thanks to cloud computing. It enables upscaling or downscaling of the services following seasonal traffic, demand, and spikes.

3.7 Education on Cloud

Cloud computing provides a new learning environment that provides an easier way to data and information for students, teachers, and researchers. The school sector is expected to receive economical, dependable, and flexible computing solutions from the cloud, allowing it to compete more successfully with larger enterprises [30]. By utilizing the cloud, students can access school tasks remotely via internet connection, teachers can rapidly submit lesson plans, and administrators can effortlessly communicate while spending less money on data storage [31].

4 Cloud Computing Adoption Factors

Technology-organizational-environment (TOE) is an approach used to describe and present the processes and factors of implementing and adopting technological advancements such as cloud computing [32].

4.1 Technological Factors

Complexity is identified as the level to which technology is recognized as quite complicated to use and understand. Cloud computing firms and companies would be hesitant to implement cloud computing systems if it feels complex. It may take them longer to understand and implement such a system. The following areas of implementation would be hampered by the complexity of cloud computing in this situation: multiple services must be continuously checked, managed, and supported [33].

Compatibility issue is another factor in adopting innovation, the level at which technological advancement is fitted to the existing technological framework, current processes, and practices. Also, the existing information resources are considered in figuring out the decision in the cloud computing adoption of an organization. In the study of Matias and Hernandez entitled "Cloud Computing Adoption Intention by MSMEs", business and cloud computing concerns and compatibility do influence the adaptation of using cloud computing in the said enterprise [34].

4.2 Organizational Factors

The caliber of its human resources, the intricacy of its organizational framework, and the size of the firm are the factors that fall under the organizational group. These general factors help with developing a beneficial ambiance and back-up with resources for the integration of innovations such as cloud computing.

Top management is also mentioned as a factor in the said adoption. Top management refers to the managerial level group that is created to support the integration of innovation in terms of its potential and cost while planning and easing management issues involved in the integration [35].

4.3 Environmental Factors

Trading partners and competitive pressure are considered great determinants in any IT adoption. As other companies invest in new innovations or high-tech companies arise, other firms face pressure to also adopt and follow the said competitors [36]. The main advantages of firms that adapt to cloud computing are improved awareness of market visibility, better data collection, and enhanced operational efficiency.

5 Cloud Computing Impacts

5.1 Agriculture

Cloud Computing Technology is a great investment for the agricultural sector since its impact affects agricultural e-commerce platforms. First, it provides an internationally centralized agricultural product marketing channel. It is also used for data mining and consumer analysis. The second is building a top-notch cloud computing approach for monitoring. It offers precise connectivity control, savvy decision-making assistance, and on-time information exchange. Third, cloud computing can determine the market demand. Agricultural sectors will know information about the latest target consumer needs. Finally, it can lower expenses and boost productivity. It also facilitated clients' internet purchases of agricultural goods. It reduces marketing costs and improves marketing efficiency [37].

5.2 Economy

Cloud computing services provide huge and widespread economic savings. It enhances the affordability of computing power and infrastructure of third-world countries. Time spent in designing, developing, acquiring, and maintaining suitable equipment, acquiring, and training human resources is significantly reduced because of cloud computing [38].

5.3 Industry

Gaining productivity, creating new services, managing supplies effectively, implementing business-to-business e-commerce, and developing a competent staff are some of the advantages of CC [39]. Moreover, CC, is a main technological foundations of Industry 4.0. It has the capability to store information in a form that supports and enables production processes to communicate with suppliers and customers in an easier way [40]. The AI industry can execute its products with access to the affordable computational power that cloud computing can provide. The significant increase in cloud computing adoption contributes to one of the factors of the AI market, with 2019 to 2026 as the expected upsurge period [41].

5.4 Labor

As a pillar of Industry 4.0, cloud computing gave new opportunities for employees that have the qualifications and skills which are related to new technologies. Jobs in these industries require both technological and human skills. It is an opportunity for more efficient workplaces, and greater profit [42]. Numerous studies have shown that as smart, autonomous programmed devices increasingly permeate the workplace, the employment rate is at risk. According to a recent study by labor economists, replacing 1,000 workers with robots decreases the worker-to-population ratio by 0.18 to 0.34 percent while increasing earnings by 0.25 to 0.5 percent [43].

5.5 Society

Cloud computing also helps the health sector. Physiologic and pathologic indicators, advanced monitoring, cloud-based health-care information systems, and prescription administration are just a few of the traditional processes and systems that ICTs offer improvements for. ICTs also created new methodologies, practices, and applications in the field of healthcare [44].

Cloud computing provides students and teachers with more flexibility as learning shifts from face-to-face to online setting. It enables schools to access resources in real-time. The recent year's disruptions demonstrated the importance of adaptable technology.

The prime concept of cloud computing is to provide ICT-enabled service systems and additional devices that are being connected to the Internet. This is why issues such as antitrust, privacy, security, jurisdiction, liability, and industrial promotion policies developed while this technology is being implemented nationally and regionally. Many existing regulations are the consequence of political agreements reached between interested parties, mediated by political and/or bureaucratic cooperation. The emergence of cloud computing as the global economy's new infrastructure will resurface and modify critical concerns that will have an impact on the world economy for the foreseeable future [45].

6 Policy Recommendations

Philippines is still a newbie to cloud computing technology in Asia as compared to its neighboring countries. The country may learn from these top ten countries about legislation and the best practices for cloud computing implementations [46].

There is no one-size-fits-all cloud computing policy. Programs are instead in place to encourage cloud providers to satisfy quality and security criteria, therefore enhancing their market position. In Germany, a non-profit organization now led the Certified Cloud initiative, which began as a government subsidy program, provides certification and a platform for 'certified' cloud services through their Trusted Cloud Portal. IT and data quality, transparency, security, data protection, and service contracts are among the certification requirements which is aimed at both cloud consumers and providers, but particularly at SMEs.

Additionally, the government may establish no broad cross-industry and cross-sector laws for cloud computing; instead, enterprises in specific sectors must comply with industry-specific legal criteria. To repeat, these nations lack general guidelines, which is why enterprises or agencies must adhere to certain sectors relating to their cloud computing industry [47]. This no cross-industry or cross-sector legislation will allow industries to fast-track the migration and intensify their implementation of cloud computing as there will be no external cross-sector interference to be considered.

Finally, the migration from Cloud First to Cloud Smart. Cloud Smart focuses primarily on policies and initiatives that enhance three pillars of cloud computing: security, procurement, and workforce. To increase security protocols, agencies must prioritize security at the data layer itself to have a multi-layer defensive approach referred to as defense-in-depth [48]. The Cloud Smart policy focuses on the workforce among agencies. In the process of migration to cloud agencies must identify possible skill gaps to be equipped.

7 Conclusion

Cloud computing provides an economic, dependable, scalable way to run applications, store data, process data, and host a web. From grid computing and client-server model, cloud computing continuously develops. The global and local trends of CC increase in terms of market size, service capability, and data center investments. Institutional, technological, and ecological factors have an impact on the embrace of cloud computing. Furthermore, CC also impacts agriculture, economy, industry, labor, and society sectors. CC affects the relationship between the producers and consumers in terms of its improvement in e-commerce in agriculture, the economy, and industry.

As the country faces different challenges and barriers in adopting cloud computing technology, agencies are encouraged to re-analyze policies. The recommendations given such as giving emphasis to programs, industry-specific legal requirements, systematic cross-border data transmission of personal data, strengthening security, procurement, and workforce as a smart implementation of cloud computing; are still appropriate to the Philippine context, although most of them are executed under Federal management.

Acknowledgments. This is to extend thankfulness to our respective families, and the Polytechnic University of the Philippines, most specially the Electronics Engineering Department for all their support and encouragement in helping us begin and complete this paper. Finally, we want to express our gratefulness to God in giving us wisdom and fortitude necessary to carry out this study and investigate the realm of cloud computing.

References

1. Wu, J., Ping, L., Ge, X., Wang, Y., Fu, J.: Cloud storage as the infrastructure of cloud computing. In: 2010 International Conference on Intelligent Computing and Cognitive Informatics, pp. 380–383. IEEE (2010)
2. Lakshminarayanan, R., Kumar, B., Raju, M.: Cloud computing benefits for educational institutions. In: Second International Conference of the Omani Society for Educational Technology. Subjects: Computers and Society (cs.CY). arXiv:1305.2616, arXiv:1305.2616v1, https://doi.org/10.48550/arXiv.1305.2616
3. Kim, W.: Cloud computing architecture. Int. J. Web Grid Serv. **9**(3), 287–303 (2013). https://doi.org/10.1504/IJWGS.2013.055724
4. Jadeja, Y., Modi, K.: Cloud computing—concepts, architecture and challenges. In: 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), pp. 877–880. IEEE (2012)
5. Varghese, B., Buyya, R.: Next-generation cloud computing: new trends and research directions. Future Gener. Comput. Syst. **79**(3), 849–861 (2018)
6. Uschie, M.: History and evolution of cloud computing: what to expect in 2025. In: Seasia Infotech (2018). https://www.seasiainfotech.com/blog/history-and-evolution-cloud-computing/
7. Gill, S.S., et al.: Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: evolution, vision, trends and open challenges. Internet Things 8, 100118 (2019)
8. Flynn, M.J.: Very high-speed computing systems. In: Proc. IEEE **54**(12), 1901–1909 (1966)
9. Fernando, C.: The Evolution of Distributed Systems (2018). https://medium.com/microservices-learning/the-evolution-of-distributed-systems-fec4d35beffd
10. Morgan, S.: The world will store 200 zettabytes of data by 2025. In: Cybercrime Magazine, Cybersecurity Ventures (2020). https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/
11. Sumina, V.: 26 cloud computing statistics, facts & trends for 2022. In: Cloudwards (2022). https://www.cloudwards.net/cloud-computing-statistics/
12. Cloud Computing Market Size, Share & Trends Analysis Report. 2022 – 2030 (2022)
13. Marr, B.: The 5 biggest cloud computing trends in 2022. In: Forbes (2021). https://www.forbes.com/sites/bernardmarr/2021/10/25/the-5-biggest-cloud-computing-trends-in-2022/?sh=6b67a40f2267
14. Donthu, N., Gustafsson, A.: Effects of COVID-19 on business and research. J. Bus. Res. **117**, 284–289 (2020)
15. Arizton Advisory and Intelligence: The Philippines Data Center Market to Attract $635 Million Investment by 2027. YCO Cloud Centers, Space DC, and Digital Edge are New Entrants—Arizton. https://www.globenewswire.com/news-release/2022/03/31/2414381/0/en/The-Philippines-Data-Center-Market-to-Attract-635-Million-Investment-by-2027-YCO-Cloud-Centers-Space-DC-and-Digital-Edge-are-New-Entrants-Arizton.html
16. de los Reyes, J.: The rise of data centers in the Philippines. In: JLL Philippines (2022). https://www.jll.com.ph/en/trends-and-insights/research/the-rise-of-data-centers-in-the-philippines

17. Hani, A.: The Philippines adopting cloud technology for business flexibility. In: OpenGov Asia (2021). https://opengovasia.com/the-philippines-adopting-cloud-technology-for-business-flexibility/
18. Schnabel, C.: PH small businesses lure cloud computing giants. In: Rappler.com (2015). https://www.rappler.com/business/industries/105310-cloud-smes-microsoft-enterprise/
19. Lagasca, V.: Cloud computing in the Philippines. In: Highland Cloud Products and Services. http://highlandcloudservices.com/cloud-computing-in-the-philippines/
20. Conoza, A.P.B.: Broader potentials in the Philippine data center market. In: Business World (2022). https://www.bworldonline.com/special-features/2022/03/28/438476/broader-potentials-in-the-philippine-data-center-market/
21. Is the Philippines ready to be APAC's next Data Center Hub? In: Manila Bulletin (2022). https://mb.com.ph/2022/02/26/is-the-philippines-ready-to-be-apacs-next-data-center-hub/. Accessed 4 Jul 2022
22. Cloud Readiness Index (CRI) of the Philippines in 2020, by segment. In: Statista (2021). https://www.statista.com/statistics/992459/philippines-cloud-readiness-index-by-segment/
23. Lugtu Jr., R.C.: Cloud computing in PH ripe for take-off. Manila Bull. **583**(27), B2 (2021). https://animorepository.dlsu.edu.ph/faculty_research/4615 (or) https://mb.com.ph/2021/07/27/cloud-computing-in-ph-ripe-for-take-off/. Accessed 4 Jul 2022
24. Odun-Ayo, I., Ajayi, O., Akanle, B., Ahuja, R.: An overview of data storage in cloud computing. In: 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), pp. 29–34. IEEE (2017)
25. Sharma, K., Singh, K.R.: Online data back-up and disaster recovery techniques in cloud computing: a review. Int. J. Eng. Innovative Technol. **2**(5), 249–254 (2012)
26. Purcell, B.M.: Big data using cloud computing. J. Technol. Res. **5**, 1–8 (2014)
27. Odun-Ayo, I., Ananya, M., Agono, F., Goddy-Worlu, R.: Cloud computing architecture: a critical analysis. In: 2018 18th International Conference on Computational Science and Applications (ICCSA), pp. 1–7. IEEE (2018)
28. Bedi, A., Pandey, N., Khatri, S.K.: Analysis of detection and prevention of malware in cloud computing environment. In: 2019 Amity International Conference on Artificial Intelligence (AICAI), pp. 918–921. IEEE (2019)
29. Reasons: Why Cloud Computing in E-Commerce is Important? https://www.newgenapps.com/blogs/importance-of-cloud-computing-in-e-commerce/. Accessed 4 Jul 2022
30. Nayar, K.B., Kumar, V.: Cost benefit analysis of cloud computing in education. Int. J. Bus. Inf. Syst. **27**(2), 205–221 (2018)
31. Gottsegen, G.: Cloud computing & education. In: Built In (2019)
32. Tornatzky, L.G., Fleischer, M., Chakrabarti, A.K.: The Processes of Technological Innovation. Lexington Books, Lexington, MA (1990)
33. Tittel, E.: Managing the complexities of cloud computing. In: CompTIA. https://www.comptia.org/blog/managing-the-complexities-of-cloud-computing. Accessed 17 Jun 2022
34. Matias, J.B., Hernandez, A.A.: Cloud computing adoption intention by MSMEsin the Philippines. Glob. Bus. Rev. 22(3), 612–633 (2021)
35. Al-Jabri, I.M., Alabdulhadi, M.H.: Factors affecting cloud computing adoption: perspectives of IT professionals. Int. J. Bus. Inf. Syst. **23**(4), 389–405 (2016)
36. Low, C., Chen, Y., Wu, M.: Understanding the determinants of cloud computing adoption. Ind. Manag. Data Syst. **111**(7), 1006–1023 (2011)
37. Zhang, C., Rao, W.: Impact of cloud computing on agricultural product e-commerce. IOP Conf. Ser. Mater. Sci. Eng. **750**(012210), 1–5 (2020)
38. Yamin, M.: Cloud economy of developing countries. World J. Soc. Sci. **3**(3), 132–142 (2013)
39. Kshetri, N.: Cloud computing in the global south: drivers, effects and policy measures. Third World Q. **32**(6), 997–1014 (2011)

40. Velásquez, N., Estévez, E.C., Pesado, P.M.: Cloud computing, bigdata and the industry 4.0 reference architectures. J. Comput. Sci. Technol. 18(3), 258–266 (2018)
41. Rosales, M.A., Magsumbol, J.-a.V., Palconit, M.G.B., Culaba, A.B., Dadios, E.P.: Artificial intelligence: the technology adoption and impact in the Philippines. In: 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), pp. 1–6. IEEE (2020)
42. Grodek-Szostak, Z., Siguencia, L.O., Szelag-Sikora, A., Marzano, G.: The impact of industry 4.0 on the labor market. In: 2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), pp. 1–5. IEEE (2020)
43. Rainie, L., Anderson, J.: The Future of Jobs and Jobs Training. Pew Research Center: Internet, Science and Techology (2020)
44. Aceto, G., Persico, V., Pescapé, A.: Industry 4.0 and health: internet of things, big data, and cloud computing for healthcare 4.0. J. Ind. Inf. Integr. **18**, 100129 (2020)
45. Kushida, K.E., Murray, J., Zysman, J.: Cloud computing: from scarcity to abundance. J. Ind. Compet. Trade **15**, 5–19 (2015). https://doi.org/10.1007/s10842-014-0188-y. Accessed 10 Jun 2022
46. Vailshery, L.: Cloud computing policy environment—country ranking 2018. In: Statista. https://www.statista.com/statistics/830031/worldwide-cloud-computing-policy-environment-country-ranking/. Accessed 4 Jul 2022
47. Zentner, L., Bensinger, V.: Q&A: cloud computing law in Germany. In: Lexology (n.d.). https://www.lexology.com/library/detail.aspx?g=36e13a7f-2df3-4687-b558-a863a9e1a56a. Accessed 4 Jul 2022
48. Kent, S.: From cloud first to cloud smart. In: United States Office of Management and Budget Official Website (n.d.). https://cloud.cio.gov/strategy/4. Accessed 4 Jul 2022

Revolutionizing the Creative Process: Exploring the Benefits and Challenges of AI-Driven Art

Sunanda Rani[1], Dong Jining[1], Dhaneshwar Shah[1](✉), Siyanda Xaba[1], and Prabhat Ranjan Singh[2]

[1] School of Art Design, Wuhan University of Technology, Wuhan, China
dhaneshwar2016@whut.edu.cn
[2] Amity Institute of Information Technology, Amity University, Patna, Bihar, India

Abstract. This paper examines how Artificial intelligence (AI) has revolutionized the creative process by enabling machines to create art and explore new mediums, technologies, and techniques in artistic expression. The research methodology employed includes a literature review, interviews with leading art professionals, and case studies. The results suggest that AI can be used to expand human creativity by allowing the artist to explore new techniques and technologies as well as to create unique works of art. However, ethical and practical considerations must be considered when working with this type of technology, such as the cost of creating AI-driven art and the imperfect output that can be generated by algorithms. Additionally, AI-driven art may not be accepted into major institutions as its medium, meaning that consideration must be taken to ensure that AI art is presented professionally and respectfully. This paper pursues to provide insight into the impact of AI technology on the creative process and explore the potential benefits and challenges associated with using AI to create artwork.

Keywords: Artificial intelligence · Deep learning · Generative adversarial networks · Digital technology · Computer vision · Digital art

1 Introduction

Artificial intelligence (AI) has been revolutionary in transforming the way art is created and expressed. With the ability to explore new mediums, technologies, and techniques, AI has enabled machines to generate unique works of art on their own. By utilizing AI technology, the creative process is no longer limited to the traditional methods of production, as machines are now capable of exploring and expressing ideas with unparalleled efficiency and accuracy [1]. Furthermore, AI can open up possibilities for the human artist to experiment with new techniques, materials, and forms of expression, while also providing a wealth of information and inspiration. As such, it is clear that AI has revolutionized the creative process by allowing machines to create art and explore new mediums, technologies, and techniques in artistic expression [2]. Additionally, AI can also be used to analyze existing works of art and gain a deeper understanding of what makes them successful or unsuccessful. This type of analysis has been especially

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 234–243, 2024.
https://doi.org/10.1007/978-3-031-50151-7_23

beneficial for digital artists, as it can help to inform the creative process and reduce the time and effort required to produce artwork [3]. Finally, AI can be used to automate certain parts of the creative process, allowing for faster and more efficient production of artwork. All of these elements contribute to an increased understanding of the potential of artificial intelligence to revolutionize the creative process.

This paper examines how AI has revolutionized the creative process by allowing machines to create art and explore new mediums, technologies, and techniques in artistic expression. Furthermore, this paper will consider ethical and practical considerations when working with AI technology, such as the cost of creating AI-driven art and the potential for imperfect output from algorithms, as well as the implications of AI-driven art being accepted into major institutions [4]. Through this research, it is hoped that a better understanding of the implications of AI on the creative process will be gained, and the potential for AI to revolutionize the art world will be highlighted. In addition to examining the potential benefits of using AI to create art, this paper also considers ethical and practical considerations that must be considered when working with this type of technology. These include the costs associated with creating AI-driven art, as well as the potential for imperfect output from algorithms. Additionally, consideration must be taken to ensure that AI-driven art is presented professionally and respectfully to be accepted into major institutions. Furthermore, complex decisions must be made surrounding who owns the rights to artwork generated by AI algorithms and how those rights are protected. As such, a thorough understanding of the implications and considerations surrounding AI-driven art is necessary to ensure a successful creative process.

2 Research Methodology

AI is an interdisciplinary field of study that seeks to understand and replicate human intelligence by creating algorithms and machines that can perceive, reason, learn, and make decisions. In recent years, AI has grown in popularity and is increasingly being used to create artwork. AI-generated artwork typically begins with a dataset or set of instructions that allows the AI to generate art [5]. AI algorithms are then used to analyze the data or instructions and generate the desired artwork. Figure 1 illustrates the various research methods that are used in AI to create artwork. The first step of the research methodology involves a literature review, where relevant research papers and other sources of knowledge related to the topic of AI-generated artwork are consulted. This step allows researchers to get an overview of existing research on the topic and identify any gaps in our knowledge base. Additionally, interviews with leading artistic professionals are conducted to gain insight into their experiences and opinions on the use of AI in the creative process. Through this step, researchers can gain a better understanding of how AI can enhance or impede the creative process. Finally, case studies are performed to evaluate the impact of AI on the creative process in real world settings. By applying these research methods, researchers can uncover the overall benefits and challenges associated with using AI to create artwork.

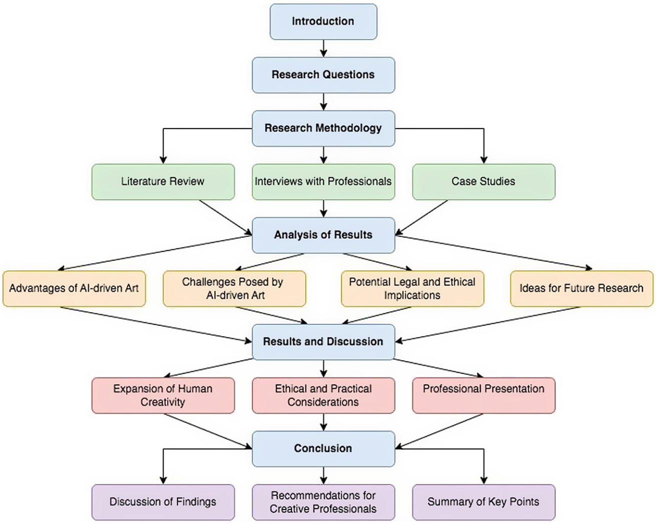


Fig. 1. Research flow diagram

2.1 Literature Review

The emergence of AI technology has had a profound effect on the creative process and revolutionized many industries. This literature review will explore the various ways in which AI has changed the creative process and how it has both benefited and hindered creativity and productivity [6]. It will look at how AI is used to create art, how AI-driven algorithms are used to produce creative works, and how AI-generated artificial creativity is impacting art and entertainment. One of the most significant changes to the creative process brought about by AI technology has been the automation of certain artistic processes, including image recognition and object recognition. The use of AI-driven computer vision techniques has enabled machines to recognize objects, interpret images, and generate artwork without human intervention. AI has also opened up new opportunities for creative expression. AI algorithms and machine learning (ML) are being used to produce music, manage sound synthesis, and generate virtual worlds with graphical designs. AI-driven algorithms can recognize and reproduce complex forms of artistic expression, from simple shapes to intricate textures. AI's ability to recognize complex patterns and use them to create artwork can help artists discover new ways of thinking and creating.

AI-based tools can assist with the creative process by providing suggestions and helping to speed up certain aspects of it, such as asset creation and content curation [7]. AI is also powering the development of tools that can assist in the creation of videos, animations, and artwork. AI-driven software has been used to optimize the workflow of creative professionals and enable them to quickly produce high-quality content. AI's ability to speed up the creative process has created both opportunities and challenges for creative professionals. While more efficient tools can give artists more freedom to experiment and explore, it is important to note that AI-driven tools are not meant to completely replace human creativity. AI should be seen as an aid rather than a substitute for creative professionals. There are several potential drawbacks associated with the use of AI-driven tools, such as the lack of flexibility and the potential to produce low-quality artwork. Additionally, it is difficult to predict how AI-generated content will affect the market or how artists will respond to this technology. AI technology is also subject to bias and ethical considerations, which should be considered when using AI for creative purposes [8]. AI is revolutionizing the creative process by automating certain aspects of the creative process and enabling new opportunities for creative expression. As AI technology continues to evolve, it is essential to consider both the potential benefits and challenges associated with its use in the creative space. To maximize its potential, creative professionals need to learn how to use AI-driven tools and understand the impacts of AI-generated content. Through this understanding, creative professionals can take advantage of the opportunities AI offers and minimize its potential disadvantages. To make sure that AI is used responsibly and ethically, it is also necessary to consider the legal and economic implications of using AI-driven tools in creative processes. Additionally, there needs to be more research on how creative professionals can use AI technology to enhance rather than replace their creative processes [9]. Overall, the emergence of AI technology has had a profound impact on the creative process and provides both opportunities and challenges for creative professionals. By understanding the implications of using AI-driven tools in the creative process, creative professionals can take advantage of the opportunities AI offers and minimize its potential disadvantages. In addition to understanding the implications of using AI-driven tools, it is also necessary to consider the legal and ethical aspects of using AI in creative processes. Finally, it is important to understand the potential impact of AI-generated content on the creative industry, as well as the benefits and challenges associated with these new forms of creativity. AI technology is changing the way creative professionals produce content and manage the creative process. As such, creative professionals need to understand the implications of using AI-driven tools to make sure that AI is used responsibly and ethically and to ensure that they are taking full advantage of the opportunities AI offers.

2.2 Interviews with Professionals

To gain a better understanding of the potential benefits and challenges posed by AI-driven art, interviews will be conducted with professional artists and other creative professionals who have already used AI-driven art in their projects. Through this research, some of the key questions asked will include how the creative process has changed and improved with the introduction of AI and what challenges have been confronted. These discussions will provide invaluable insights into the real-world experiences and impact of AI-driven art on the creative profession and offer a more nuanced look at the implications of this technology.

AI driven art has revolutionized the creative process, providing unprecedented speed and accuracy in the creation of projects. Artists have personally experienced the benefits of utilizing it in their projects, such as being able to experiment with different styles quickly and easily. Despite some challenges, such as the steep learning curve and spotty reliability, artists have seen great success in its use. AI-driven art has opened up many new possibilities in the creative profession, from interactive experiences to increased precision and speed. Artists have taken steps to ensure ethical use of AI-driven art, and have found it to be an invaluable tool in their creative practice [10]. Artists' advice to others looking to incorporate AI-driven art into their creative process is to do their research and remain up to date on the latest developments in the field. They should also consider how their work might be impacted by AI-driven art, both positively and negatively, and ensure that their use of it is ethical and responsible. With careful consideration, AI-driven art can open up new opportunities to express one's creativity while staying mindful of its potential pitfalls. AI driven art technologies, as well as to understand the potential challenges and ethical considerations that may arise. Ultimately, AI-driven art promises to vastly improve how we create projects and can provide new avenues for exploring and expressing creativity. By leveraging AI-driven art, professionals in the creative industry will be able to unlock new and exciting possibilities for their projects, resulting in a more dynamic and engaging experience for their audiences. Additionally, it is important to remain mindful of ethical considerations when utilizing AI-driven art, and to stay up to date on the latest developments in the field [11]. With the right approach and by learning how to properly integrate AI-driven art into the creative process, professionals will be able to take their projects to the next level. By understanding the potential of AI-driven art and properly implementing it into their creative practice, professionals in the creative industry can experience the full power of this technology and unlock new possibilities for their projects. With AI-driven art, professionals can create projects faster, with greater accuracy and deeper levels of personalization. Additionally, by expanding the scope of their creative projects, professionals can make them even more engaging and interactive for their audiences. Ultimately, AI-driven art has the potential to revolutionize the way that professionals in the creative industry approach their projects. By embracing technology and learning how to use it properly, professionals can leverage the power of AI-driven art to create unique and engaging projects. With the right approach, AI-driven art can be an invaluable tool for creating powerful and meaningful projects that stand out from the crowd. By staying up to date on the latest developments in the field and understanding how to properly integrate it into the creation process, professionals will be able to take their projects to the next level.

2.3 Case Study

For this research project, case studies will be conducted to further investigate the potential use of AI-driven art for creative processes. Through these case studies, information will be gathered about the successes and failures of using AI-driven art for particular projects. This information will be used to provide a better understanding of how AI-driven art can improve creative processes, as well as its limitations. The case studies will involve looking at projects that have already implemented AI-driven art, as well as experiments and research conducted on potential uses for the technology. The following are a few case studies that can be used for the research paper:

Sougwen Chung is a renowned artist who uses artificial intelligence (AI) to create interactive artwork, exploring automation, creative expression, and ethical considerations. Her pieces "Project Atman" and "Counter-point" feature AI-generated abstract art and robots painting together. David Li's "Blob Opera" is another revolutionary example of AI-driven art, which allows users to craft tunes and melodies without prior music knowledge or ability [12]. By utilizing AI-driven tools, the system can generate songs tailored to a user's creative input, opening up infinite possibilities for musicians, composers, and artists alike. The implications of AI-driven art are already being seen in the world of music and culture, changing traditional barriers of entry in the process. The Mila Study, an AI-based research project headed by Dr. Yoshua Bengio at the Université de Montréal's MILA lab, studied the potential of using images from simulated 3D environments and domain adaptation tasks to create more creative and realistic AI-driven art. Through machine learning, the team was able to develop models that can recognize objects, apply materials and textures, and understand how textures should move with an object's movements. Simultaneously, British artist Matt Talbot created an interactive art installation, Cacophonic Choir, which uses machine learning algorithms and AI technology to create dynamic soundscapes in response to movement. This project raises important questions about the potential for AI in the creative process and demonstrates its ability. Both of these projects highlight the growing potential of AI in the creative process and demonstrate how it can revolutionize the way art is created. With the Mila Study providing a deeper understanding of AI-driven art and automation of the creative process, and Cacophonic Choir exploring how AI can be used for interactive artwork, artists and content creators have more options with which to create unique, high-quality works. Portrait of Edmond de Belamy is an artwork created by AI. It was generated using a GANs technique and sold for $432,500, making it one of the most expensive pieces of AI-generated artwork ever sold. The portrait was composed of an algorithm that incorporates the style of 15th and 16th century portraits to create something new [13]. This artwork has sparked debate in the art world regarding the creative implications of AI, with some arguing that it would devalue the field and others exploring its potential. AI-driven art has become increasingly popular and accepted in mainstream art, encouraging artists to explore machine learning algorithms as part of their creative process. This has sparked debates around the ethics of using artificial intelligence in the creative process and what its implications might be for the future of art.

3 Results and Discussion

3.1 Expansion of Human Creativity

Driven art has the potential to revolutionize the creative process by allowing creators to explore new possibilities in their artwork. In the past, emerging technologies such as computer graphics and digital audio have enabled artists to push the boundaries of creativity, creating entirely new forms of media. AI-driven art adds to this vibrant creative sphere by providing meaningful insights into what is possible with creative tools and techniques.AI can unlock human creativity by analyzing an artist's work and offering suggestions for improvements or new directions based on what the AI learns about the artist's style and preferences [14]. AI can also find correlations between a creator's work and that of other successful artists, enabling the artist to build off of the most successful elements from their predecessors. Additionally, AI can offer insights into how artwork can appeal to different audiences, giving creators a better understanding of how best to reach their target market. By allowing artists to incorporate AI into their creative process, AI-driven art has the potential to open doors to new levels of innovation and creativity. AI has a particularly promising impact on visual arts and art forms that rely heavily on computationally intensive techniques, such as 3D modeling or video editing. With AI, artists can explore more complex visuals than ever before and create artwork that stands out from the competition [15]. Additionally, AI-driven art can be used to generate artworks that can reflect the artist's thoughts and feelings in a whole new way, offering a range of possibilities for creating emotionally resonant works. Despite these potential benefits, there are also challenges to incorporating AI into the creative process. Creators must carefully consider how to incorporate AI into their work without it overshadowing their own creativity or supplanting their artistic vision. Additionally, AI-driven art may require a deep investment of time and resources to create works of quality, meaning creators may have to adjust their expectations regarding the scale and scope of their projects. Finally, creators may find themselves facing ethical dilemmas when creating artwork with AI, as they must make sure that the outputs reflect their own values and beliefs, not those of the AI. Despite these challenges, AI-driven art has the potential to revolutionize the creative process by opening new doors for creativity and innovation. By allowing creators to leverage the insights and tools provided by AI, it is possible to create complex and emotionally resonant artwork that stands out from the competition. As creators become more comfortable with incorporating AI into their creative processes, the creative possibilities offered by AI-driven art will only expand, leading to a new era of exciting and imaginative artwork.

3.2 Ethical and Practical Considerations

AI-driven art has revolutionized the creative process and made it more accessible to all. However, ethical and practical considerations should be considered when utilizing AI in art. Ethically, artists should be mindful of copyright, authorship, and ownership issues regarding AI-generated art. Additionally, consent must generally be obtained from any subjects used in the artwork, as well as ensuring the artwork is labeled appropriately to inform the public of its origin. Practically, questions must be addressed such as how the artwork is going to be sold and displayed, as well as how it can be properly valued and monetized [16]. AI-generated art also raises legal issues, such as intellectual property regulations and copyright legislation. To protect the artist and the viewer, appropriate protocols should be implemented. At the same time, it is important to recognize the potential of AI-driven art. Its use can greatly expand the creative possibilities for both amateur and professional artists, as well as offer a wide range of opportunities for education and research. To ensure that this potential is realized responsibly, artists must ensure that ethical and practical considerations are considered when using AI in their artwork. By adhering to these standards, AI-driven art can revolutionize the creative process while preserving the safety and security of both the artist and the viewer.

3.3 Professional Presentation

AI revolutionizing the field of art. AI can enable artists to create complex work and take advantage of new tools, such as virtual reality and augmented reality. The potential of AI-driven art to revolutionize how creativity is expressed is immense and can provide a platform for innovation and faster production of works. However, stakeholders must come together to ensure equitable and secure system that respects individual rights. The future of AI in art looks bright and can truly revolutionize the creative process [17]. In order to ensure the successful implementation of AI-driven art, it is important to consider the following aspects during the development and use of such technologies. First, careful consideration must be given to legal frameworks, to ensure that the rights of all parties involved are protected. Additionally, technical considerations need to be made in order to ensure that the system is secure and trustworthy. Finally, ethical implications and potential unintended consequences should also be discussed and considered, as these may have an impact on the wider creative community. By taking these considerations into account, AI-driven art could empower both professional and amateur artists, leading to more creative, innovative and unique works of art. AI has the potential to revolutionize the creative process, offering new possibilities for both artistic expression and techniques. For example, AI-driven tools can help artists create new forms of art, such as generative art, which is created algorithmically using sets of rules and machine learning algorithms. AI-driven tools can also enable artists to generate new ideas faster by automating certain tasks, such as pattern recognition or data analysis. Furthermore, AI-driven tools could help to eliminate tedious manual processes and help to reduce the cost of production. Additionally, AI may offer artists a way to utilize data sets that can provide insights that they may not be able to discern on their own. By utilizing AI, artists can gain new perspectives on their work and identify trends that may allow them to explore untapped creative potential or provide a way to access previously unreachable and unexplored creative territories. The results of this could include works of art with a previously unseen level of complexity and depth. However, there could be challenges and risks associated with the exploration of AI-driven art, including questions surrounding privacy, ethics, and copyright. It is therefore important for professionals to understand both the potential benefits and challenges associated with AI-driven art in order to effectively present, discuss, and debate these topics.

4 Conclusion

The potential of AI to revolutionize the field of art is immense and could result in new opportunities for artists, art collectors and art galleries to explore uncharted territory. AI has been used in interactive art installations, providing interesting and immersive experiences for viewers. AI can also enable artists to create complex work that may have seemed impossible before due to the level of precision that can be attained using these tools. Furthermore, AI can enable artists to take advantage of new tools, such as virtual reality and augmented reality, to create artwork that takes advantage of such technologies. There are numerous benefits to such a system, including increased speed, improved accuracy, and greater efficiency. However, there are also some challenges to be aware of, such as safety risks and copyright issues. To ensure that all those involved benefit from this technology, policy makers, software developers, and artists must work together to create an equitable and secure system that respects individual creative rights and encourages collaboration, creativity, and innovation. In conclusion, AI-driven art has the potential to revolutionize how creativity is expressed and provide a platform for innovation and faster production of works. With everyone involved on the same page, the future of AI-driven art looks bright and can truly revolutionize the creative process.

Continuous dialogue between stakeholders will be crucial in order to keep up with the ever-evolving nature. The future of AI in art is bright and could revolutionize the way creativity is expressed and produced. Furthermore, this technology has the potential to create interactive art installations that provide interesting and immersive experiences for viewers. There is no doubt that the future of AI-driven art is bright and full of potential.

References

1. de Leon, J.A., et al.: Deep learning approach to 2d capacitive resistivity imaging inversion. In: Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569, pp. 459–468. Springer, Cham (2023)
2. Historillo, J.F., Rosales, M., Merlin, M., Mandia, E.B.: Color and image analysis approach in determination of soluble copper in water using tannic reaction analysis. In: Intelligent Computing and Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569, pp. 958–971. Springer, Cham (2023)
3. Hossain, M.A., Hasan, M.A.F.M.R.: Activity identification from natural images using deep CNN. In: Intelligent Computing and Optimization. ICO 2020. Advances in Intelligent Systems and Computing, vol. 1324, pp. 693–707. Springer, Cham (2021)
4. Wei, X., Wei, R.: Research on interactive art design system based on artificial intelligence technology. Smart Innov. Syst. Technol. **298**, 159–166 (2022)
5. Li, L.: The impact of artificial intelligence painting on contemporary art from disco diffusion's painting creation experiment. In: 2022 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML), pp. 52-56. IEEE (2022)
6. Thomas, J.J., Pillai, N.: A deep learning framework on generation of image descriptions with bidirectional recurrent neural networks. In: Intelligent Computing and Optimization. ICO 2018. Advances in Intelligent Systems and Computing, vol. 866, pp. 219–230. Springer, Cham (2019)
7. Hitsuwari, J., Ueda, Y., Yun, W., Nomura, M.: Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. Comput. Human Behav. **139**, 107502 (2023)
8. Hertzmann, A.: Visual indeterminacy in GAN art. Leonardo **53**(4), 424–428 (2020)
9. Leiser, A., Schlippe, T.: AI in art: simulating the human painting process. In: ArtsIT, Interactivity and Game Creation. ArtsIT 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 422, pp. 295–308. Springer, Cham (2022)
10. Hung, M.C., Nakatsu, R., Tosa, N., Kusumi, T.: Learning of art style using AI and its evaluation based on psychological experiments. Int. J. Arts Technol. **14**(3), 171–191 (2022)
11. Guljajeva, V., Sola, M.C.: Dream painter: an interactive art installation bridging audience interaction, robotics, and creative AI. In: MM'22: Proceedings of the 30th ACM International Conference on Multimedia, pp. 7235–7236 (2022)
12. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: a review. IEEE Trans. Vis. Comput. Graph. **26**(11), 3365–3385 (2020)
13. Goenaga, M.A.: A critique of contemporary artificial intelligence art: who is Edmond de Belamy?. AusArt J. Res. Art. **8**(1), 49–64 (2020)
14. Chen, W., Shidujaman, M., Xuelin, T.: AiArt: towards artificial intelligence art. In: MMEDIA 2020: The Twelfth International Conference on Advances in Multimedia, pp. 47–52 (2020)
15. Hedes, B.-C.: Happening, a controversial hybrid way of cultural expression. Stud. Univ. Babes-Bolyai Dramatica **67**(2), 123–140 (2022)

16. Blomfield, I., Lenette, C.: Artistic representations of refugees: what is the role of the artist? J. Intercult. Stud. **39**(3), 322–338 (2018)
17. Liu, Z.: Research and implementation of digital art media system based on big data aesthetics. Smart Innov. Syst. Technol. **298**, 195–203 (2022)

Vehicle License Plates Recognition Using Object Detection and Classification Algorithms

Nguyen Tan Cam[1,2](✉), Dong Thi Ngoc Tram[1,2], and Duy Nguyen[1,2]

[1] Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City, Vietnam
camnt@uit.edu.vn, tram.dong04@gmail.com, duynt.14@grad.uit.edu.vn

[2] Vietnam National University, Ho Chi Minh City, Vietnam

Abstract. To better manage vehicles in traffic, license plates are one of the best ways to identify vehicles. Each number plate is unique for a vehicle with distinct characteristics of vehicle type, color and information management such as: year of manufacture, origin, registration date, inspection date, weight, owner. The license plate is also the basis for the management and statistics of vehicles involved in transporting goods, transporting passengers, importing and exporting goods, residential areas, etc. However, license plates have may be partially or completely obscured. This causes misidentifications if we only use license plate recognition solutions. In this study, we propose a method of vehicle identification based on object recognition in obscured environments. The experimental results from this study show that our proposed system has many advantages in practice.

Keywords: Object recognition · Deep learning · Obscured environments · License plate recognition · YOLO · SSD

1 Introduction

Because of the exponential growth of the transportation industry, the need for efficient and accurate vehicle identification has become increasingly important. License plate recognition (LPR) systems can be used in various applications, such as traffic control, parking management, toll collection, and law enforcement. LPR systems utilize computer vision techniques to recognize license plates of vehicles from images or videos captured by cameras mounted on roads, buildings, or police vehicles.

The recognition of license plates is a challenging problem due to the variability in plate designs, lighting conditions, and occlusion. To address these challenges, recent research has focused on applying object recognition and classification algorithms to LPR systems. Object recognition algorithms are used to detect license plates within an image or video frame, while classification algorithms are used to recognize the characters on the plate.

This paper presents a novel approach that uses object recognition and classification algorithms for the recognition of vehicle license plates. The proposed method utilizes

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 244–252, 2024.
https://doi.org/10.1007/978-3-031-50151-7_24

a deep learning framework to accurately detect license plates via object detection algorithms. The effectiveness of the proposed aproach is evaluated on a large dataset in the wild, and the results demonstrate superior performance compared to existing methods.

The paper has five main sections. The Sect. 1 is the introduction. The Sect. 2 is related works. Some studies related to LPR systems are discussed in this section. Section 3 describes the proposed approach that can be used to recognize the license plates. The experimental results are analyzed in Sect. 4. Finally, Sect. 5 concludes the paper and discusses potential future directions for research in this field.

2 Related Works

Currently there are many studies that are used to determine the license plates of vehicles [1–10]. Chen and Wang [1] proposed the novel framework for license plate recognition that combines broad learning system and convolutional networks. They proposed pixel-level based binary classification by using the fully convolutional network is designed for pixel-level two-class classification. They employ a trained AdaBoost cascade classifier for character segmentation. With Macau license plates demonstrate their experiment results show that their proposed approach outperforms several state-of-the-art approaches. However, this study will also encounter difficulties in recognizing number plates in obscured environments.

Chen [2] proposed the novel approach that can be used for automatic license plates recognition (ALPR). It using YOLO (You Only Look Once) framework to detect car license plates. Their focus is on recognizing Taiwan's car license plates. They use 7 convolutional layers of YOLO to detect a single class. They use the AOLP dataset. They test the system under different scenario, including rainy backgrounds, dimness and darkness, different saturation and hues of images. Their results demonstrate the robustness and effectiveness of our proposed approach for ALPR in challenging conditions. However, this study will also encounter difficulties in recognizing number plates in obscured environments.

Weihong and Jiaoyang [3] explored the use of deep learning to recognize the license plates. They focus on addressing the three challenges: license plate skew, license plate blur, and image noise. They classified deep learning algorithms with not only indirect detection algorithms but also direct detection algorithms. They analyze the strengths and weaknesses of current approaches in license plates detection. Additionally, they compared different license plate recognition systems in terms of data sets, accuracy, workstation, and time. They also compared and evaluated existing public datasets based on number of pictures, the resolution. Finally, they provide a perspective on future research directions in license plate recognition.

3 Proposed System

In this study, the vehicle identification system based on object recognition algorithms is presented as shown in Fig. 1.

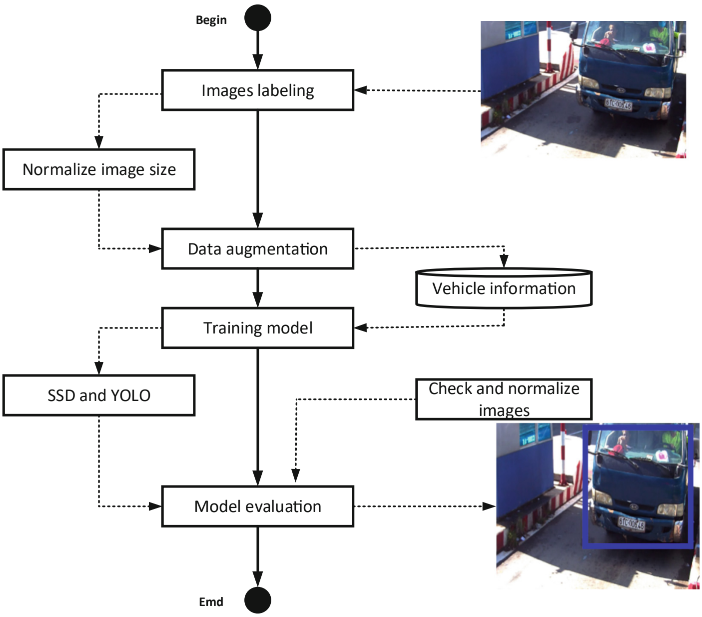


Fig. 1. The proposed system's architecture

3.1 Images Labeling

In this study, we use the LabelIng tool [11] to perform the images labeling process. Figure 2 shows the results of labeling of vehicles with ID 51D04183.

Fig. 2. The results of labeling of vehicles with ID 51D04183

Figure 3 shows the vehicle parameters after labeling. In which, (1): belongs to class = 0; (2): the horizontal center coordinates/width of the figure = 0.390972; (3): the center coordinate of the height/height of the figure = 0.486979; (4): frame width/image width = 0.645833; (5): frame height/figure height = 0.939236.

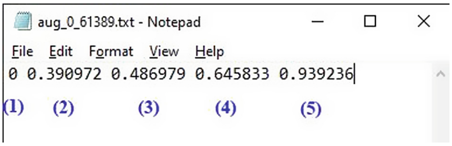


Fig. 3. The vehicle parameters after labeling

3.2 Data Augmentation

In this study, we use ImageDataGenerator library of the Keras [12] to perform data augmentation. The actual image of the vehicle with ID 51D04183 is shown in Fig. 4.

Fig. 4. An example of the actual image of the vehicle with ID 51D04183

Figure 5 presents the augmented image. This is a case of Loss of frame detail with 02 squares with a size of 10% of the frame. Besides, the algorithm increases frame brightness by 25%. Also reduces frame brightness by 25%. In addition, the algorithm blurs the frame by 03 pixels, causing as many frames as 03% per pixel.

Fig. 5. An example of the augmented images

3.3 Training Model

In this study, we use SSD [13] and YOLO [14] to perform object detection and classification. SSD architecture based on CNN for feature detection is performed in two phases. The first phase is used to extract feature maps. The second phase applies convolution filters to identify features. The SSD uses VGG16 to extract the feature map. Then detect objects using VGG16's Conv4_3 class. Each prediction has not only a bounding box but also 21 points for each class. If there is no object, class "0" is added. The class with the highest score is selected as the class for the detected object. The bounding box is closed. This SSD is suitable for real-time object recognition applications.

YOLO was introduced in 2015 with the ability to handle 45 frames per second (FPS), much better than the Faster R-CNN model that only reached 7 FPS. There are many versions of YOLO released and each version shows a significant improvement (especially in terms of speed). YOLO in object detection means "You only look once". That is, we only need to look once to be able to detect the object. Although YOLO is not the best model, it is the fastest object detection models. YOLO can achieve near-real-time speeds. YOLO can be used to detect multi objects in the same image.

4 Evaluation

4.1 Dataset

Each vehicle retrieves 50 images. In which, 30% is used as a training base, 10% is a test basis during training, and 10% is used as a verification basis after training.

In this study, we use data augmentation algorithm to increase the number of samples. The operations shift left, shift right, shift up, shift down. Besides, we increase the brightness, decrease the brightness. In addition, we add noise into samples. These tasks are used to simulate various real-world environmental conditions.

In this paper, we use the labelImg.py toolkit [15] for labeling. Data used to train the recognition model is showd in Table 1.

Table 1. The parameters of the dataset

Parameter	Values
Resize	512 (W) $\times$ 409 (H), 96 dpi, 24 bit
Training data	100 classes: 3,000 images
Validation data	100 class: 1,000 images
Predict data	100 classes: 1,000 images, 720 (W) $\times$ 576 (H), 96 dpi, 24 bits (for each layer 10 images)

Table 2. The test scenario configuration

Parameter	SSD	YOLO
System	Colaboratory	Colaboratory
Library	TensorFlow	TensorFlow
Class number	100	100
Size	512 (W) × 409 (H), 96 dpi, 24 bit	512 (W) × 409 (H), 96 dpi, 24 bit
Training data	3,000/100 class	3,000/100 class
Validation data	1,000/100 class	1,000/100 class
Predict data	1,000/100 class	1,000/100 class
Pre-train model	ssd_mobilenet_v2	darknet53.conv.74 (https://pjreddie.com/media/files/darknet53.conv.74)
Epochs number	30,000	30,000

4.2 Evaluation

The test scenario configuration is described in Table 2.

Table 3 shows the evaluation results. The training time is 45 h 35 min. The mAP (Mean Average Precision) reached 96.30% at 30,000 batches when we use YOLO.

The training time of SSD is 55 h 05 min. The mAP of SSD reached 77.93% at 30,000 batches.

Table 3. The evaluation results

Model	mAP (%)	Training time
SSD	77.93	55 h 05 min
YOLO	96.30	45 h 35 min

Figure 6 shows the test results when using the YOLO algorithm to identify vehicles.

Figure 7 shows the test results when using the SSD algorithm to identify vehicles.

From the evaluation results and the comparison table (Table 3), the YOLO model gives the best results. Therefore, this paper recommends choosing the YOLO model to apply to the vehicle license plate information prediction system.

5 Conclusion and Future Works

After testing the deep learning algorithms on the image dataset, the YOLO model achieved positive results and can be applied to the vehicle license plate prediction model. The paper has also achieved the following research objectives: Built an automatic vehicle identification system, extracted vehicle license plate information, built a standard data

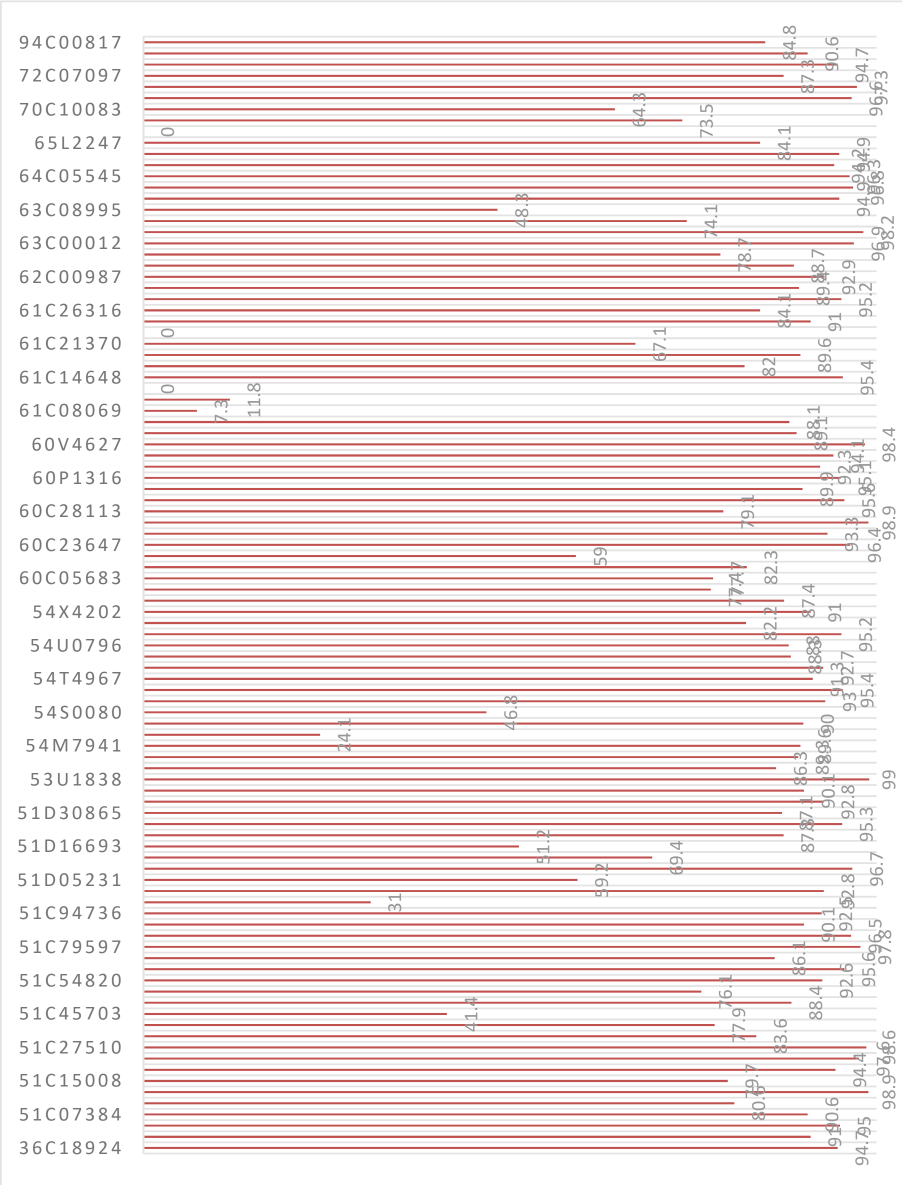


Fig. 6. The test results when using the YOLO algorithm

set with 100 vehicles participating in regular and continuous market activities with 5,000 pictures for each vehicle (including 3,000 images as training data and 1,000 images as control data). The paper analyzed and evaluated the dataset of the most engaged vehicles, learning several predictive deep learning algorithms that fit the dataset. The paper has

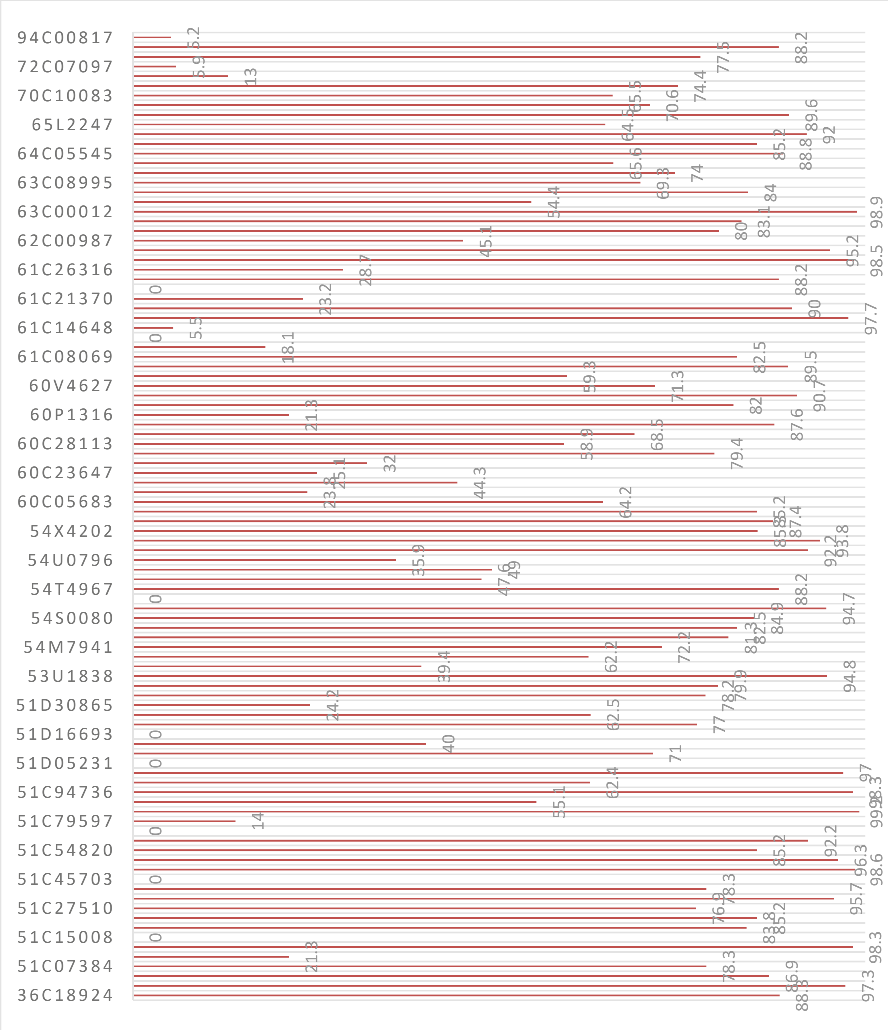


Fig. 7. The test results when using the SSD algorithm

successfully applied machine deep algorithms in prediction. Thereby, this paper has chosen the YOLO algorithm for the predictive model of each vehicle. The algorithm runs on a dataset of 100 different vehicles and gives high accuracy, improving the performance of vehicle license plate information recognition.

Acknowledgement. This research was supported by The VNUHCM-University of Information Technology's Scientific Research Support Fund.

References

1. Chen, C.L.P., Wang, B.: Random-positioned license plate recognition using hybrid broad learning system and convolutional networks. IEEE Trans. Intell. Transp. Syst. **23**, 444–456 (2022)
2. Chen, R.-C.: Automatic license plate recognition via sliding-window darknet-YOLO deep learning. Image Vis. Comput. **87**, 47–56 (2019)
3. Weihong, W., Jiaoyang, T.: Research on license plate recognition algorithms based on deep learning in complex environment. IEEE Access **8**, 91661–91675 (2020)
4. Duan, N., Cui, J., Liu, L., Zheng, L.: An end to end recognition for license plates using convolutional neural networks. IEEE Intell. Transp. Syst. Mag. **13**, 177–188 (2019)
5. Saidani, T., Touati, Y.E.: A vehicle plate recognition system based on deep learning algorithms. Multimedia Tools Appl. **80**(30), 36237–36248 (2021). https://doi.org/10.1007/s11042-021-11233-z
6. Joshua, Hendryli, J., Herwindiati, D.E.: Automatic license plate recognition for parking system using convolutional neural networks. In: 2020 International Conference on Information Management and Technology (ICIMTech), pp. 71–74. IEEE (2020)
7. Wang, W., Yang, J., Chen, M., Wang, P.: A light CNN for end-to-end car license plates detection and recognition. IEEE Access **7**, 173875–173883 (2019)
8. Riaz, W., Azeem, A., Chenqiang, G., Yuxi, Z., Saifullah, Khalid, W.: YOLO based recognition method for automatic license plate recognition. In: 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 87–90. IEEE (2020)
9. Sun, H., Fu, M., Abdussalam, A., Huang, Z., Sun, S., Wang, W.: License plate detection and recognition based on the YOLO detector and CRNN-12. In: Signal and Information Processing, Networking and Computers: Proceedings of the 4th International Conference on Signal and Information Processing, Networking and Computers (ICSINC). Lecture Notes in Electrical Engineering, vol. 494, pp. 66–74. Springer, Singapore (2019)
10. Sharma, V., Jain, M., Jain, T., Mishra, R.: License plate detection and recognition using OpenCV–Python. In: Recent Innovations in Computing. Lecture Notes in Electrical Engineering, vol. 832, pp. 251–261. Springer, Singapore (2022)
11. Tzutalin: Label Studio Is a Modern, Multi-Modal Data Annotation Tool (2022). https://github.com/heartexlabs/labelImg
12. Chollet, F.: Keras: Simple. Flexible. Powerful (2022). https://keras.io/
13. Liu, W., et al.: SSD: single shot multibox detector. In: Computer Vision—ECCV 2016. Lecture Notes in Computer Science, vol. 9905, pp. 21–37. Springer, Cham (2016)
14. Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B.: A review of yolo algorithm developments. Procedia Comput. Sci. **199**, 1066–1073 (2022)
15. Lin, T.: labelImg 1.4.0 (2023). https://pypi.org/project/labelImg/1.4.0/

The Impact of HIV/AIDS in the Workplace and Its Effects on an Organization's Productivity: A Case of the Aluminium Industry in South Africa

Xolanani Ntombela[1], Gill Manion[1], Elias Munapo[2](✉), Martin Chanza[2], and Kolentino Mpeta[2]

[1] Graduate School of Business and Leadership, University of KwaZulu-Natal, KwaZulu-Natal, RSA

[2] Department of Business Statistics and Operations Research, North-West University, Potchefstroom, RSA

Elias.Munapo@nwu.ac.za

Abstract. The paper presents a case study of empirical realities, attitudes, and behavior of workers about the impact of HIV/AIDS at an aluminium company in South Africa, accompanied by initiatives to examine and evaluate the response of the company's management to HIV/AIDS in the workplace. The study also explored the company's HIV/AIDS program concerning the efficiency and effective implementation of existing workplace program/s for both affected and unaffected workers. It recommended possible actions to the existing HIV/AIDS program in the company. What emerged from the study is that the Health Unit, Clinic and company management have taken severe steps to challenge the realities of HIV/AIDS in the workplace. According to the questionnaires given out, it seems that most of the respondents are aware of the main factors contributing to the spread of HIV/AIDS. The staff members are aware of how condoms can act as a barrier and prevent the disease from spreading. The results also showed that there is still some confusion among the workforce regarding some important concerns related to understanding the virus and the disease. It is recommended that the company's Human Resources and Industrial Relations departments, along with the Health Unit in addition to support by the trade union operating in the company, expand and diversify their use of new and innovative communication tactics.

Keywords: HIV · AIDS · Impact · Triangulation method · Workplace

1 Introduction

HIV stands for Human immunodeficiency virus (HIV), the virus causing Acquired immunodeficiency syndrome (AIDS). AIDS attacks and destroys the body's immune system leading to immunodeficiency, thus leaving infected individuals unable to fight off opportunistic diseases.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 253–269, 2024.
https://doi.org/10.1007/978-3-031-50151-7_25

HIV/AIDS epidemic has had, and continues to have a severe effect on the world of work [1]. It affects economic growth and social development due to reduced supply of labour, loss of skilled and experienced workers, absenteeism, early retirement, increased labour costs for employers resulting from health insurance, reduced productivity, contracting tax base in the economy and loss of family income thus aggravating poverty [2].

The HIV/AIDS epidemic significantly impacts the South African working class and the economy as it affects the workplace, thus causing operational disruptions in organizations. This disruption worsens when more qualified and experienced employees are absent since finding a replacement is difficult.

As a result, the HIV/AIDS menace has become the center of research at the workplace in both the developed and developing worlds [3]. Most of the results of these studies conducted so far are not readily available to the public. In addition, there is limited understanding of the factors that determine whether employees are willing to have an HIV test at the workplace or not.

Previous studies on voluntary testing and counseling (VCT) about the implications of HIV infection and how to avoid transmitting the virus have indicated that the convenience of getting the test and the attitude of the staff administering it are all very important [4].

The purpose of this paper was to investigate the impact of HIV/AIDS at the aluminium company in South Africa and the objectives were:

- to investigate the impact of HIV/AIDS on the workforce,
- to investigate the company's HIV/AIDS intervention programmes,
- to find out employees' perceptions and knowledge of HIV/AIDS and
- to recommend a company-wide HIV/AIDS intervention programme.

The following sub-sections are presented: literature review, methods, results, discussions, conclusions, and recommendations.

2 Literature Review

HIV is highly stigmatized worldwide [5]. It is a social process that is used to legitimize discrimination. It is the prejudice and discrimination targeted at people perceived to have AIDS virus as given by [6]. According to [7], stigma may take a social, physical, verbal or institutional form. Stigma usually results in discrimination. It is faced in various contexts, including household, community, workplace and healthcare settings [8]. Prejudice and discrimination is usually shown by fear of physical contact with people living with HIV, negative judgments about them, and exclusion or marginalization [9]. The prejudice is attributed to lack of information on interventions to address it and measures to monitor the progress of those interventions [10].

The fear of stigma negatively affects the way in which individuals and families care for those affected with HIV [11]. It is one of the most serious obstacles in the fight against HIV/AIDS the world over, including South Africa. This is due to the fact that HIV infection is widely perceived as an outcome of low moral character [9]. The latest findings by [12], show that stigma is still challenging in South African societies.

Besides its disastrous effects in households, stigma presents tangible challenges to both private and public sector organizations since the virus often infects people in their productive ages. Thus, if HIV prevalence reaches a high level in a firm, the impact may be heavy as workers affected are likely to be absent more often. Besides, AIDS-related deaths reduce the number of available workers in the most productive years. This leads to a firm productivity decline and, ultimately, profitability [13]. HIV/AIDS in the workplace has not become the center of research in Africa in general [3]. The fact that South Africa has the highest global prevalence of HIV/AIDS, with an estimated 6.4 million of its residents reporting having the disease in 2012 [14], lends credence to this.

Research on HIV/AIDS at the workplace is ongoing. [15] worked on HIV/AIDS workplace programs in the automotive industry. The results showed that 12% among management and 42% among the cleaners had sexual interactions with more than one partner in the previous 12 months. Even though there are massive HIV/AIDS awareness programs, more than 50% of the employees are not willing to disclose their status.

In reaffirming the country's commitment to addressing the impact of HIV/AIDS at the workplace, the South African government published legally binding guidelines in 2003 that offered employers advice on enhancing productivity and boosting the morale of infected and affected workers. According to the Guidelines [16], stigmatization and discrimination will not assist HIV-positive workers continue to contribute to the economy of their country. The report advised businesses to form workplace committees to address HIV/AIDS and gender concerns as well as create preventative health policies. The guidelines were helpful in eradicating unjust workplace discrimination, fostering a non-discriminatory attitude toward those who are HIV-positive. Furthermore, the guidelines supported practical HIV/AIDS management strategies in the workplace as well as striking a balance between each party's rights and obligations.

3 Research Methodology

In order to gather data, formulate conclusions, and offer recommendations regarding the impact of the HIV/AIDS epidemic at the company, this study used the triangulation method. Both semi-structured interviews and questionnaires were used in this methodology.

3.1 Questionnaire

To get the most responses possible from respondents, "closed" questions were utilized in this study. These questions easily lend themselves to quantitative analysis. The questions were formatted in three ways to meet the requirements of the research topic and to guarantee thorough coverage of the research objectives:

- The single-option response, which was yes or no,
- Mutliple choice responses and
- Likert scale type questions.

As the targeted company is not a medical or health organization, it would have been useless to use technical or medical terminology in the questionnaire. The questionnaire's primary goal was to open and keep open, a line of communication between the researchers and the participants. Great effort was taken to avoid using language that might lead to miscommunication or researcher bias.

3.2 Pretesting Questionnaire

[17] made the point that pretesting the questionnaire with a small sample of respondents who shared traits with the target population could help ensure accuracy and consistency of responses. During the pretest, a sample size of 4 to 30 may be utilized [17]. First, the proposed questionnaire's content validity and terminology were evaluated by testing it on coworkers before being used in the interviews. This pretesting helped identify any flaws before the instrument was given to a large number of participants verbally or through a written questionnaire, which decreased bias [18].

A validity examination of the questionnaire as a research tool is based on the pretesting of the questionnaire. A modified questionnaire was then created to test its applicability and suitability for the intended audience.

3.3 Reliability of the Research Instrument

According to [19], reliability is primarily focused on issues of a measure's consistency. This concept is comparable to that put out by [20], who described reliability as a sign of a measuring instrument's consistency or stability. In this study, the Cronbach's coefficient was found to be 0.857, which showed a very good rate of internal reliability for the questionnaire.

3.4 Semi-structured Interviews

Since it was a triangular study, the company's nursing sister participated in a semi-structured interview to ascertain the following:

- The frequency of sick leave brought on by the HIV/AIDS pandemic.
- The pandemic's toll on the mortality rate of skilled employees.
- The aluminum company's propensity for early medical boarding.

Interviews are frequently used as a source for gathering data or proof. Contrary to questionnaires, interviews typically contain the following characteristics:

Open-ended inquiries.

- A massive and perhaps limitless information yield;
- Qualitative, phenomenologically based study.

In a semi-structured interview, structured and open-ended questions are mixed together [21]. The inquiries made to the business sister were created especially for this investigation. In this interview, leading questions, double-barreled questions, and confusing questions were avoided.

Open-ended questions were employed since they provided the participant with the chance to provide in-depth responses on issues for which structured questions were deemed insufficient. Both qualitative and quantitative approaches could be incorporated in the question design.

3.5 Pretesting Interview Questions

First, the corporate doctor pretested the interview questions. This pre-testing helped the researcher refine the questions' wording, get rid of statements that were unclear, and evaluate the data-collection strategy.

3.6 Sample Size and Distribution of the Questionnaire

All South African firm employees make up the study population, which is the wider corporate population from which the study sample was recruited. The sample frame's population consisted of 2500 workers, of whom 2000 were permanent employees and 500 were contractors. For this investigation, probability sampling was chosen, which is supported by the following facts:

- The sample obtained from one of the branches of the company was fully representative of all company employees in South Africa.
- Each worker in the sampling frame had an equal chance of getting chosen. Employees of the branch comprised the sampling frame.
- Results from the sample can be extrapolated to the entire population.

As a sample frame, the head office (branch personnel) was used. 100 employees were chosen at random from the payroll list using a simple random selection method, and they received the surveys. Of these 100 employees, 80 people completed and retained questionnaires giving an 80% response rate (sample size = 80/100 = 80%). These employees had different statuses, and various employee levels and were of varying gender groups within the company.

3.7 Limitations

The researchers could only interview the company sister. She was the only one available at the company. The company management is not involved in health matters and was not supportive in interviewing employees about health-related issues. The nurse is responsible for all clinic issues.

4 Data Analysis

4.1 Age of Respondent

The Respondents were asked to indicate their age within various age bands. Figure 1 depicts the proportions of each category. The bulk of survey participants were under 36 years old.

4.2 Gender of Respondents

Respondents indicated their gender as shown in Fig. 2. It was noted that the sample was dominated by men (91%).

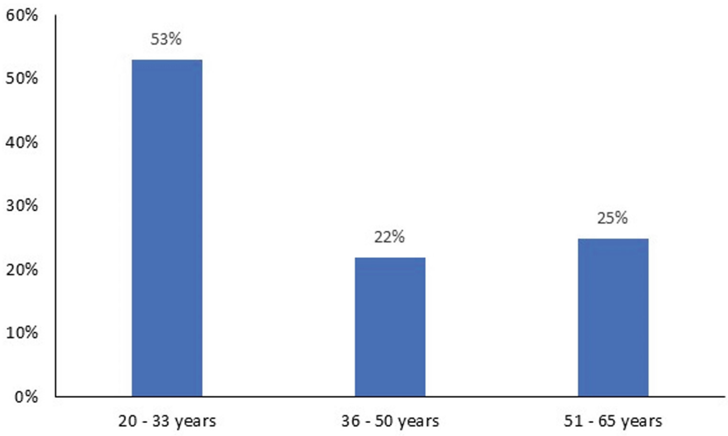


Fig. 1. Distribution of respondents by age

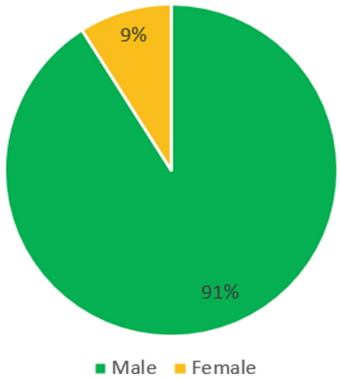


Fig. 2. Gender distribution

4.3 Educational Level of Respondents

The respondent's level of education is shown in Fig. 3. Most respondents went to the high school level.

4.4 Occupation/Job Status

Respondents indicated their job statuses as shown in Fig. 4. The majority of the respondents were skilled workers.

4.5 Years of Employment

The employment period was one of the questions in the questionnaire and the results are displayed in Table 1. It was noted that most respondents have been working for the company for more than 10 years.

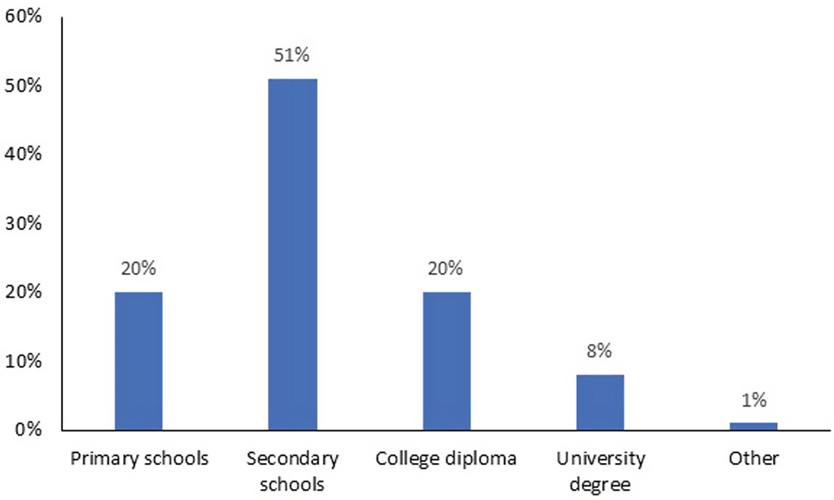


Fig. 3. Educational level

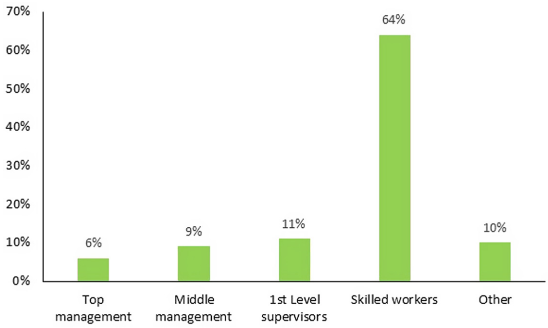


Fig. 4. Occupation/Job status

Table 1. Years of experience

Years of employment (YE)	Number of respondents (%)
YE $\leq$ 2 yrs	16
2 yrs $<$ YE $\leq$ 5 yrs	14
5 yrs $<$ YE $\leq$ 10 yrs	25
YE $>$ 10 yrs	45

4.6 Spread of HIV/AIDS

The majority of respondents identified unprotected sexual contact and blood contact with HIV-positive individuals as the primary causes of HIV/AIDS transmission. This is reflected in Table 2 and also note that there were multiple responses evident.

Table 2. Mode of HIV transmission

Mode of transmission	Number of respondents
Blood contact with an HIV-positive person	78
Sharing a cup with an HIV-positive person	4
Unprotected sexual intercourse	79
All of the above	3

In Table 3, 68 male respondents are of the opinion that condoms help in preventing HIV transmission, while 7 female respondents do not think that condoms help to prevent the spread of HIV. Most of the employees (85%) believe that condoms prevent the spread of HIV. Also, not all of the female respondents believe that condoms are useless in preventing HIV/AIDS.

Table 3. Condoms and prevention of HIV/AIDS

Gender	Affirmative	Negative
Blood contact with an HIV-positive person	68	5
Sharing a cup with an HIV-positive person	0	7
As % age of the total respondents	85	15

4.7 Sharing Utensils with HIV-Infected Employees

Table 4 shows that most of the respondents would not mind sharing utensils with HIV-infected colleagues. About contracting the disease by just working next to infected people, 14 employees believed so.

Table 4. Sharing utensils with HIV-infected employee

Mode of transmission	Number of respondents
Will share utensils with HIV infected colleagues	66
Will not share utensils with HIV infected colleagues	14

4.8 Open HIV/AIDS Discussions

Regarding the HIV/AIDS open discussions, only 3 respondents did have conversations with their colleagues at the work place as shown in Tables 5 and 6.

Table 5. Discussions of HIV/AIDS at the workplace

HIV/AIDS discussions at the workplace	Number of respondents (%)
Had discussions	96
No discussions	4

Table 6. Reasons why there are discussions of HIV/AIDS at the workplace

Reasons	Number of respondents
Because of wellness day	47
Because of HIV/AIDS day	39
The role of the clinic is vital	41
Part of safety procedures	44
HIV/AIDS is unacceptable	47
Discussions opened among us	54

4.9 Wellness Clinics

Everyone who responded said the company runs wellness clinics. The frequency of wellness clinics at the organization, as reported by employees, is shown in Fig. 5. The majority of employees said that the employer holds wellness clinics two to three times a year.

4.10 HIV Testing As Part of the Wellness Clinic Function

All respondents said that the function of the wellness clinic included HIV testing. It was made clear that staff members could arrange private HIV testing after scheduling a consultation with the nurse.

4.11 Company Assistance to Workers Who Have Tested Positive

Regarding the processes associated with the company's endeavors to assist workers identified as HIV positive, responses were tabulated as shown in Table 7.

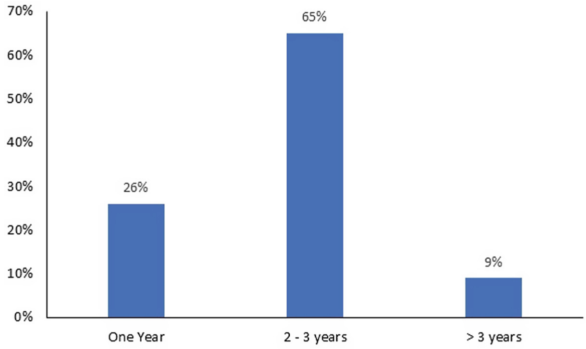


Fig. 5. Frequency of wellness clinics

Table 7. Company assistance towards HIV/ infected employees

Help from company	Number of respondents (%)
Send employees to a local clinic or hospital	45
Management does not provide any help	5
Management follows them up every six months	37
Do not know	13

Table 8. Caring for HIV/AIDS-affected people

Caring for the affected employees	Number of respondents (%)
Yes	60
No	13
Do not know	27

4.12 Caring for HIV/AIDS Affected People

Most of respondents (60%) thought that the company is caring, 13% did not think so and 27% did not know. The information is given in Table 8.

4.13 Efficient Implementation of Programs

The majority of the employees (77%) are of the opinion that the wellness program at the company is effectively implemented, 5% believe that it is not effective and 18% did not know, as shown in Table 9.

Table 9. Responses to the effectiveness of the wellness program

Wellness program	Number of respondents (%)
Yes	77
No	5
Do not know	18

4.14 Respondent Reasons for Efficient/Inefficient Implementation of Programs

4.14.1 Reasons for Efficient Implementation

The respondents gave the following reasons as the reasons for the efficient implementation of programs by the company in its fight against the HIV/AIDS pandemic:

- fewer sick workers than before,
- one sex partner campaign,
- employees are given ARVs and condoms,
- encouragement of employees to be tested,
- programs conducted frequently,
- help is readily available at the clinic,
- counseling taking place,
- other HIV/AIDS-related awareness campaigns and
- there are follow-ups to the HIV/AIDS programs at the company.

4.14.2 Reason for Inefficient Implementation

The respondents believed that management needs to put more effort into communicating with workers regarding existing and future problems related to HIV/AIDS.

4.14.3 Knowledge of Colleagues Taking Sick Leave Because of HIV/AIDS

Of the 80 respondents, 15% said that they knew 'many' colleagues who took sick leave because of HIV/AIDS-related issues, 20% of respondents indicated that there were 'not many' while 65% indicated they did not know. The responses are shown in Fig. 6.

4.15 Frequency of Colleagues' Sick Leave

On how often colleagues infected with HIV/AIDS take sick leave, 4% of respondents had the opinion that it was very often, 10% believed that it was often, 3% thought that it was not often and 83% did not know. The information is shown in Fig. 7.

4.16 Fear of HIV/AIDS

About the fear of HIV/AIDS, 74% of the respondents believe that workers fear HIV/AIDS, 9 do not believe so and 17% did not know (Table 10).

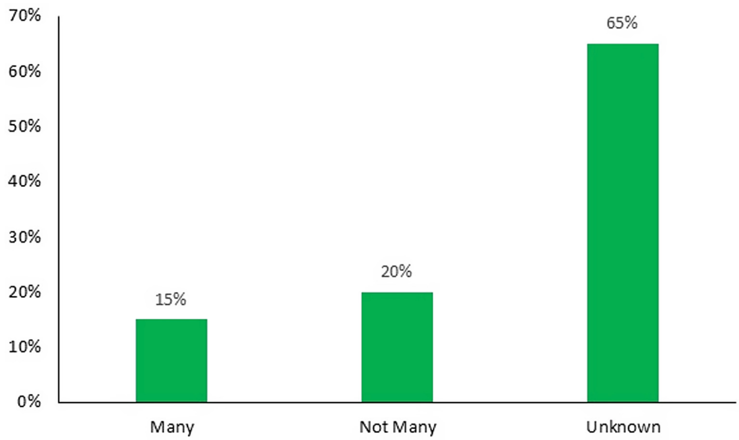


Fig. 6. Knowledge of colleagues taking sick leave because of HIV/AIDS

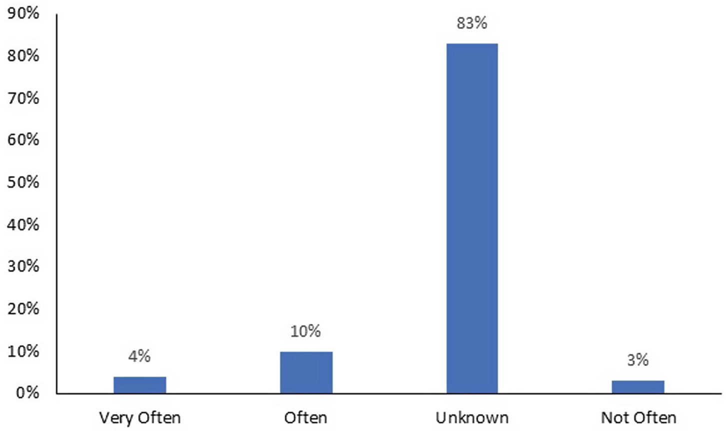


Fig. 7. Knowledge of colleagues taking sick leave because of HIV/AIDS

Table 10. Fear of HIV/AIDS

Fear of HIV/AIDS	Number of respondents (%)
Yes	74
No	9
Do not know	17

5 Discussion of Results

The investigation made clear that there is an excellent working relationship between the Clinic, the head of the Health Unit, and the Company Nurse. The corporation spends a lot of money on ARVs, and wellness clinics are held twice a year. There is also a robust plan in place for any employees who become ill due to the pandemic. However it was discovered that there are no official records outlining the effects of HIV/AIDS, which is a problem that needs to be fixed.

According to the study, the vast majority of respondents are educated about the fundamental factors that contribute to the transmission of HIV/AIDS, and a nearly equal number are aware of how condoms can act as a barrier and preventative measure against the disease.

According to the responses, there is still some ignorance among employees on such important issues related to knowledge with regard to the virus and the sickness when it comes to the use of utensils in the spread of HIV/AIDS, despite the efforts of the government and the company.

Throughout the course of the investigation, it became clear that there are open discussions among employees on the shop floor, which is contrary to earlier findings of eminent HIV/AIDS medical and social science researchers.

The explanations given for this fact on the working floor are more positive, as workers claimed that the company clinic's activities and its numerous instructional duties were crucial in these conversations. Thus, the company's proactive involvement highlights these conversations on the shop floor, demonstrating the company's adherence to the CODES OF GOOD PRACTICE of the Department of Labour SA [16], World AIDS Day, and the Wellness Days, all of which encourage meaningful conversations.

Such responses, however, are incongruent with the indication of incomplete awareness of the Wellness Day schedule, which may be due to forgetfulness, the potential of employees being on leave during the clinic's time, as well as other unidentified causes.

The company's compliance with the Code of Good Practice on HIV/AIDS issues in the workplace is demonstrated once more by the HIV/AIDS testing at the clinic, which is an essential and required service. By offering employees HIV testing, the company helps to improve the creation of a safe working environment.

The majority of the employees are well informed about the procedures involved in the business's efforts to support employees who have been diagnosed as HIV-positive and have good understanding of the company's obligations towards HIV/AIDS workers.

Given that the prevalence of HIV/AIDS among working people always has negative effects on their families and communities in terms of socioeconomic as well as psychological markers, such as a rise in poverty rates, decreased savings in the home, social exclusion, etc. According to the answers to the pertinent questions, the business would benefit from honing its communication approach in response to these problems.

Although a small percentage of respondents indicated that they are aware of many colleagues who were on sick leave due to HIV/AIDS, it is important to interpret and understand the interviewees' responses regarding their knowledge of coworkers who take leave because of the disease in the context of the stigma and fear that exist among South African's working class and poor communities in both urban and rural areas. Many of the employees expressed their concern of HIV/AIDS.

6 Recommendations and Conclusions

In the private sector, there are instances of businesses that have worked hard over the years to lessen the misery of their industrious employees and to carefully abide by the progressive laws and Codes of Good Practice that have been outlined in the context of this paper. Of course, expecting all industries and businesses in the private sector to abide by the law and rules would be utopian, but this is ultimately how the market works. It is hoped that the findings and recommendations of this research will be helpful to employers, employees, their families, and the state. Case studies like the current one have the potential to inform future policy planning as well as the company in question, and the pool of skilled workers is unquestionably small.

6.1 Recommendations

The empirical findings of the thesis lead to a number of recommendations, including:

The comments make it clear that, despite the general satisfaction with the management's strategy and actions done regarding HIV/AIDS in the workplace, communication tactics against the epidemic need to be strengthened. Hence, it became clear that such plans required coordination between the company's Human Relations and Industrial Relations departments, as well as the Health Unit. This must entail spreading messages about abstinence and the value of only having one sexual partner using a number of communication channels. The trade union bodies that are active on the work floor, such as the shop stewards and the Shop Stewards Council, must also be involved in this coordination. Due to cultural norms and beliefs, such a collaborative effort might be crucial in teaching employees who have grown accustomed to working with several partners over time.

The current anti-HIV/AIDS soap operas, such as "SOUL CITY" and "GAZLAM", which are immensely popular among the African working class, could be used in more frequent preventative programs. During lunch and tea breaks, these programs might be broadcast. Although HULLETS use this technique, no research has been done on how it affects employee behavior. Although the particular nature of them is unknown, it is known that this organization has employed similar tactics for a number of years.

The management of infected employees needs to be made more transparent. Given the delicate nature of the situation, it is crucial that the union and the company's pertinent divisions work together in a coordinated manner. Such collaboration will undoubtedly result in more favorable outcomes because it may demonstrate to the shop floor employees that both the employer and employee sections are concerned with their welfare. The peace and excellent labor relations that will result from such an initiative will also benefit the workplace.

Responses made it clear that there is a disconnect between vulnerability rates and HIV/AIDS program implementation. Therefore, it is crucial that there be better coordination between program implementation and monitoring of vulnerability rates. This may be accomplished by including a person, even on a part-time basis, in the duties and responsibilities of the Health Office. An addition like this can be advantageous since, despite the nurse's best efforts, any organizational or logistical support might prove to be really helpful. An employee from Human Resources or Industrial Relations may offer the further assistance.

The department or another individual who is qualified and knowledgeable to carry out such tasks. If such an appointment is made, it is crucial that management consult the union. Such a function may occasionally be filled by a knowledgeable and experienced shop-steward from a trade union.

It is crucial that a study be conducted to provide a scientific evaluation of the impact of HIV/AIDS on neighboring communities. This study should be funded by the company's management and conducted with union agreement and cooperation. Since many of the workers reside in these areas and others are involved in multiple-partner relationships there, this will be important in the long run. The results of such a project will support the organization of the workers and the company in developing prevention and other future initiatives.

With quick employee surveys, it is necessary to regularly monitor and evaluate the programs that the organization has already implemented. Although they have a good understanding of HIV/AIDS, there are significant gaps in their knowledge (as demonstrated in the case of attitudes towards utensil usage).

Given the inferior quality of state-sponsored condoms, there is a need for communication pertaining to the expanded condom usage and the potential for administration and the unions to reach a financial and social agreement concerning the acquisition of commercially branded condoms.

The commitment of management to current programs and the potential for their expansion must be consistently communicated to the workforce. The workers must also receive messages urging them to seek testing as well as counseling services either inside the company or outside, at clinics, hospitals, or mobile state facilities. This is crucial since many employees can feel too afraid, hesitant, or intimidated to undergo testing at the company health office for obvious reasons.

All employees must have access to a notice board that lists hospitals and clinics where all HIV/AIDS patients can receive free testing and ARVs. All District Department of Health offices in South Africa offer these boards for free.

Keeping accurate records for all HIV/AIDS patients is crucially important, and the corporation must work with the union to make this happen. This will enable all parties involved to understand the state of affairs at hand and establish appropriate short-, medium-, and long-term plans. Also, it is crucial that the business maintain thorough data regarding all HIV/AIDS patients in collaboration with the union.

6.2 Conclusions

The key problem investigated is the impact of HIV/AIDS on the shop floor, the workforce's attitudes and knowledge and the management's actions to rectify existing issues. The primary health issue of South African business sector leaders right now is HIV and AIDS, which forms the basis for this research. The effects of this epidemic on workers and business in general are getting worse every day, thus it is expected that companies will respond to HIV/AIDS in the workplace in a way that is both humane and compliant with the law and state rules. Therefore, it has been universally accepted that private sector businesses should be required to take the appropriate steps and initiatives to address a problem of this nature that has both direct and indirect adverse effects on productivity and, consequently, profitability.

6.3 Further Research

There is need to explore other possible factors contributing to the spread of the disease. These include cultural beliefs such as

(a) strengthening of charms or muthi by sleeping with:

- a pregnant woman who is not your wife,
- a woman during her menstrual period,
- a close relative or
- another man.

This is common with some African man and they believe this will work if the weird sexual act is unprotected.

(b) improving the bond between father and son or mother and daughter:

This is done by cutting the wrist blood veins of the two and pressing their arms so that blood of the parent flows in that of the child and vice versa. Parent-child bond improvements are normally done when the children are in their teens.

Even though Christianity and civilization have eaten a large chunk of the dark side of Africa, the belief that witchcraft can be used for promotion or to instill fear at the workplace is still there. These old aged cultural fallacies are not acceptable and must be studied so as to come up with effective ways to permanently keep them out of our societies.

References

1. Chatora, B., Chibanda, H., Kampata, L., Wilbroad, M.: HIV/AIDS workplace policy addressing epidemic drivers through workplace programs. BMC Public Health **18**(1), 1–14 (2018)
2. Ndobeni, Z.N.: The Effective Use of Legal Protection to Combat Stigma and Discrimination Related to HIV and AIDS in a Workplace: A Case Study in KwaZulu Natal. Magister Legum: Labour Law, Faculty of Law, Nelson Mandela University (2018)
3. Achwoka, D., Mutave, R., Oyugi, J.O., Achia, T.: Tackling an emerging epidemic: the burden of non-communicable diseases among people living with HIV/AIDS in sub-Saharan Africa. Pan Afr. Med. J. **36**(271), 1–9 (2020)
4. (NACP—INDIA) National AIDS Control Programme http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---ilo_aids/documents/legaldocument/wcms_117311.pdf. Last accessed 22 Apr 2015
5. Dong, X., et al.: HIV-related stigma and discrimination amongst healthcare providers in Guangzhou, China. BMC Public Health **18**(1), 1–10 (2018)
6. Musonda, B.: Descriptive Study of HIV/AIDS-Related Stigma Experienced by People Living with HIV/AIDS in Healthcare Settings Offering Family Planning and HIV Services at Pollen Health Centre in Kabwe, Zambia. Doctoral dissertation, Faculty of Business and Information Technology, Cavendish University (2020)

7. Dewi, D.M.S.K., Sari, J.D.E., Fatah, M.Z., Astutik, E.: Stigma and discrimination against people living with HIV and AIDS in Banyuwangi, East Java, Indonesia. In: 4th International Symposium on Health Research (ISHR 2019), pp. 154–159. Atlantis Press (2020)
8. Tran, B.X., Than, P.Q.T., Tran, T.T., Nguyen, C.T., Latkin, C.A.: Changing sources of stigma against patients with HIV/AIDS in the rapid expansion of antiretroviral treatment services in Vietnam. BioMed Res. Int. **2019**(4208638), 1–9 (2019)
9. Mahamboro, D.B., Fauk, N.K., Ward, P.R., Merry, M.S., Siri, T.A., Mwanri, L.: HIV stigma and moral judgement: qualitative exploration of the experiences of HIV stigma and discrimination among married men living with HIV in Yogyakarta. Int. J. Environ. Res. Public Health **17**(2), 636 (2020)
10. Andersson, G.Z., et al.: Stigma reduction interventions in people living with HIV to improve health-related quality of life. Lancet HIV **7**(2), e129–e140 (2020)
11. Kimera, E., et al.: Experiences and effects of HIV-related stigma among youth living with HIV/AIDS in Western Uganda: a photovoice study. PLoS ONE **15**(4), e0232359 (2020)
12. Hlongwane, N., Madiba, S.: Navigating life with HIV as an older adult in South African communities: a phenomenological study. Int. J. Environ. Res. Public Health **17**(16), 5797 (2020)
13. Faruk, M.: The impact of HIV/AIDS on service delivery in Iganga district: a case study of Buyangasub-county. A Research Dissertation Submitted to the College of Humanities and Social Sciences for the Partial Fulfillment of a Bachelor's Degree in Social Work and Social Administration at Kampala International University (2019)
14. Shisana, O., Rehle, T., Simbayi, L.C., Zuma, K., Jooste, S., Zungu, N., et al.: South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. HSRC Press, Cape Town (2014)
15. Steenkamp, L., Von der Marwitz, J., Baasner-Weihs, F., Pietersen, J.: Should HIV and AIDS workplace programmes still be advocated in the automotive industry? S. Afr. J. Hum. Res. Manage.**13**(1), 1–10 (2015). https://sajhrm.co.za/index.php/sajhrm/article/view/609/html. Last accessed 13 May 2015
16. (DL 2010) Department of Labour HIV/Aids Technical Assistance Guidelines (May 2003). Director General KZN 2010 MDG Progress Report to Cabinet, 13 Oct 2010
17. Hair, J., Babin, B., Money, A.H., Samonel, P.: Essentials of Business Research Methods. Wiley, Hoboken, America (2005)
18. Sekaran, U.: Research Methods for Business, 3rd edn. Wiley, New Jersey, U.S.A (1999)
19. Vaske, J.J., Beaman, J., Sponarski, C.C.: Rethinking internal consistency in Cronbach's alpha. Leis. Sci. **39**(2), 163–173 (2017)
20. Taber, K.S.: The use of Cronbach's alpha when developing and reporting research instruments in science education. Res. Sci. Educ. **48**, 1273–1296 (2018)
21. Roulston, K., Choi, M.: Qualitative interviews. In: The SAGE Handbook of Qualitative Data Collection, pp. 233–249. Sage Publications Ltd. (2018)

Inhibition Performance of Generic Compounds as Corrosion Inhibitor in Combined Chloride and Carbonation Environment

Ashish Kumar Tiwari[1](✉), Sunil Garhwal[2], Anil Garhwal[3], and Anirban Banik[4]

[1] Department of Civil Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India
tiwariashish841@gmail.com
[2] Department of Civil Engineering, Maharishi Markandeshwar (Deemed to Be University), Mullana, Ambala, Haryana 133203, India
[3] School of Architecture and Planning, Banasthali Vidyapith, Tonk, Rajasthan 304022, India
[4] Department of Civil Engineering, National Institute of Technology Sikkim, RavanglaSikkim, South Sikkim 737139, India

Abstract. Corrosion inhibitors are the most commonly used preventive measure because of their beneficial effect like ease of use, effectiveness, cost and eco-friendly nature. This study aims to examine the inhibition efficiency of some generic compounds selected on the basis of their functional groups in combined chloride and carbonation environment. Results obtain concludes that combined environment creates vulnerable situation of rebar. Presence of generic compounds effectively inhibit corrosion mechanism and reduces the corrosion activity by acting as a barrier against corrosive ions. Functional group like amine and carboxylic can effectively use to reduce the combined effect of corrosion.

Keywords: Corrosion · Generic compounds · Functional groups · Corrosion inhibition

1 Introduction

Durability of reinforced concrete (RC) structure is the serious problem in today era. The term durability not only affects the integrity of RC structures but also leads to change in service life of RC. Among the various durability issues, corrosion is the serious threat as it directly challenges the structure performance and may cause complete deterioration in some years only (say about 5 to 10 years). Also, it leads to huge economic loss due to maintenance and demolition process every year (around 3.4% of world GDP) [1]. Chloride ions and carbon-dioxide are the two major causes that spark corrosion of embedded steel and considerable attention has been given by the research community from the last few decades to understand the corrosion process accurately. Although, individual action of chloride and carbonation effect the integrity of RC structure by decreasing its service life but when they attack in combination, the effect can be more severe and complex. Carbonation of concrete not only reduces the pH of concrete pore solution but also

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 270–278, 2024.
https://doi.org/10.1007/978-3-031-50151-7_26

cause lots of microstructural changes. Along with this, effect of carbonation on chloride redistribution may cause a negative impact for steel by increasing the availability more number of corrosive ions near the vicinity of rebar [2–8].

There are various preventive techniques available to increase the service life of concrete structure, but application of corrosion inhibitors (CI) is proved to be most economical and effective measure to reduce corrosion initiation. In the past, various commercial and generic chemicals have been used to mitigate the corrosion process in chloride or carbonation environment [9–18], but the studies are limited to the exploration of CI for the individual environment only. The severity of combined action can assumed to be high as both uniform corrosion (due to carbonation) and pitting (due to chloride) can occur simultaneously.

Due to variations in their molecular structure, corrosion inhibitors (CI) can be broadly divided into two categories: inorganic and organic. Inorganic CIs consist of nitrate (NO^{3-}), nitrite ($NO_2{}^{-}$), chromate ($CrO_4{}^{2-}$), and molybdate ($MoO_4{}^{2-}$) in their molecular structure, while organic CIs contain amines (NH2), carboxyl (–COOH), hydroxyl (OH), and thiol (SH) groups. The advantage of using organic compounds as CIs is that they are non-toxic. The effectiveness of organic inhibitors depends on their structure [19]. Amines (which contain a basic nitrogen atom as a lone pair electron) and aldehydes are considered as adsorption inhibitors (chemisorption or physisorption). Carboxylic acids have polar groups, where oxygen contains lone electron pairs that can bond chemically and physically on the surface of rebar. They form carboxylate anions, which can absorb onto the steel surface. In contrast to amine groups, carboxylic groups act as proton donors and are considered Brønsted-Lowry acids [20]. In amino alcohol-based organic inhibitors, the main component is an amino-alcohol, which is the volatile component and transported mainly by gas diffusion. The second component is generally an acid component. This acid component is reported to react with hydration products [21]. Reaction with calcium hydroxide results in a gel formation that can block the pores of the concrete [22] and reduces the penetration of chloride. Amine and ester-based inhibitors have dual actions in concrete, as the amine compound acts as an inhibitor while the carboxylate ester compound has a pore-blocking effect, which blocks the ingress of chlorides [23]. The inhibition mechanism of carboxyl groups, amine and amino acid groups was also studied by Ormellese et al. (2011) [24] in a chloride-rich environment and found that amino acids showed some inhibition effect, and carboxylate substances showed very good inhibition by electron attractor or electro donor effect or electrostatic effect.

The idea of our research is to incorporate an environmental-friendly and non-toxic corrosion inhibitor based on amine and carboxylate group for the efficient corrosion protection. The chemicals were selected upon their molecular structure and the efforts have been made to get the diversity in the structure during selection process. The objective of present study is to investigate the effectiveness of corrosion inhibitor under combined chloride and carbonation environment.

2 Research Significance

Selection of suitable compound to mitigate corrosion of reinforced steel is the crucial step. Earlier the compounds used as corrosion inhibitor was mainly tested either in chloride contaminated environment or against carbonation induced corrosion. In combined environment, study on the efficiency of different functional group is quiet scanty. The result presented in this study could be able to give the clear vision in defining the effectiveness of different functional group in most aggressive environment.

3 Experimental Program

3.1 Materials

QST steel rebar specimens were used as test specimen in the present study. Table 1 shows the chemical composition of specimen used in the study. Test specimen was prepared by taking 60 mm length of 12 mm diameter QST rebar, rubbed by wire brush and sand paper to remove any rust present on the surface. After removing any rust product, specimen were washed with analytical grade hexane and allowed to air dry for some time. After air-drying, the length of the specimen were coated with two layer of epoxy leaving bottom 3.65 cm^2 area for testing. Also, each specimen was drill and threaded about 10 mm length and 4 mm diameter from the top. This was done to provide the electrical connection between the specimen and electrochemical setup by means of nut and bolts arrangement during testing. The cross-sectional details of the specimen are shown in Fig. 1.

Table 1. Chemical composition of steel used

Element	Percent (%)
C	0.3
S	0.055
P	0.055
N	120 (ppm)
Fe	Bal

To mimic the different exposure conditions (i.e. uncontaminated concrete environment and combined environment), concrete solutions were prepared. For preparation of pore solution representing uncontaminated concrete environment, distilled waste was admixed with calcium hydroxide ($Ca(OH)_2$) to get 0.035M solution (consider as fully saturated solution). The solution was mixed properly and left undisturbed for 24 h in sealed beaker to avoid any atmospheric contact. After 24 h, the solution was filtered and considered as interstitial concrete pore solution (Sol. 1). The pH of the Sol. 1 was measure by pH meter accurately. The initial pH of the prepared solution was 13. For simulation of combined chloride and carbonation environment, the filtered solution was carbonated by bubbling the Sol. 1 with pure CO_2 till pH drops to 8. Passing CO_2 from the solution cause formation of carbonic acid (H_2CO_3) and reaction with present calcium (Ca) from the solution it forms calcium carbonate ($CaCO_3$) precipitate. The precipitate was filtered out by using whatmann filter paper. The filtered solution was mixed with 3.5% w/w NaCl to introduce chloride in the solution and the final solution mimicking the combined environment is named as Sol. 2. All the above mentioned process was carried out at room temperature (around 25°C).

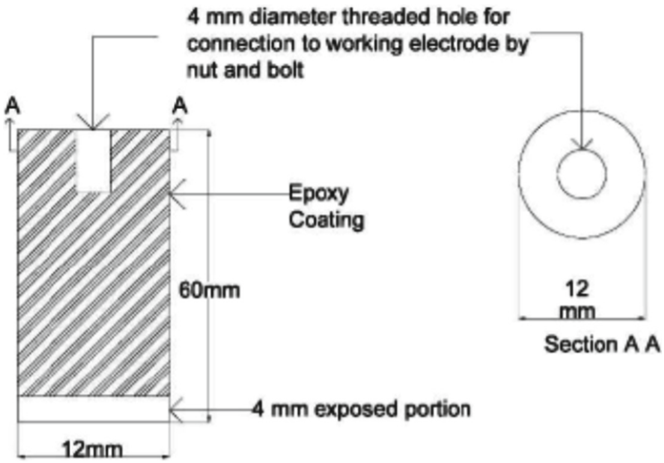


Fig. 1. Details of rebar cross-section

The chemicals were selected upon their molecular structure and presence of functional group. The molecular structure of the used chemical as corrosion inhibitor in the present study are shown in Fig. 2. The chemicals were added in Sol. 2 at a concentration 1% w/w. The nomenclature and pH details of different solution are presented in Table 2.

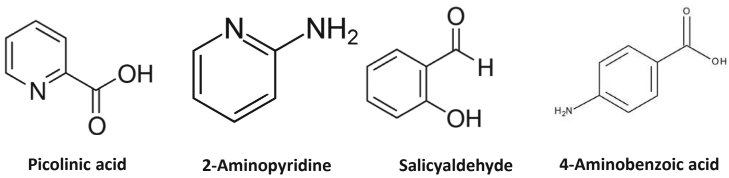


Fig. 2. Molecular structure of generic compounds used as corrosion inhibitor

3.2 Testing

The specimens were immersed in their respective solutions for 240 h before testing in order to assess the long-term effectiveness of the generic compounds used. To conduct the test, a three-electrode setup was employed, consisting of a steel specimen as the working

Table 2. Details of solution used

Abbreviation	Detail	pH
Sol. 1	Distilled water + $Ca(OH)_2$ (0.035M)	13
Sol. 2	Sol. 1 + CO_2 bubbled + NaCl (3.5%)	8
Sol. 3	Sol. 2 + Picolinic acid (PA)	7.5
Sol. 4	Sol.2 + 2aminopyridine (2AP)	8.2
Sol. 5	Sol.2 + salicyaldehyde (Sac)	6.5
Sol. 6	Sol. 2 + 4-aminobenzoic acid (AbA)	8.1

electrode, a saturated calomel electrode as the reference electrode, and platinum as the counter electrode. Potentiodynamic polarization (PDP) scanning was carried out at a scan rate of 60 mV/minute over a potential range of − 250 to 1500 mV with an offset from the rest potential. Prior to scanning, the specimen's potential was stabilized to reach its open circuit potential (OCP) to ensure accurate results. The tip of the calomel electrode was placed near the exposed area to eliminate the effect of IR drop. Tafel extrapolation was performed on the polarization curves to obtain the corrosion current densities (i_{corr}) for each solution. The PDP curves were recorded using an ACM field machine (serial no 1463), and triplicate specimens were tested to ensure the reliability of the data. The specimens were tested at a temperature of 25 ± 2°C.

4 Results

The potentiodynamic behaviour of different solution with immersion durations are shown in Figs. 3 and 4. Figure 3 shows the anodic polarization curve for Sol. 1 and Sol. 2 after 1 h or immersion. Test was done after 1 h of immersion is to study the passivation behaviour of test specimen in uncontaminated and contaminated environment.

From the Fig. 3 it can be seen three well defined zones are visible that for Sol. 1 i.e. active zone, passive zone and transpassive or pitting zone. Active zone represents the diffusion process of iron (Fe) whereas passive zone represents the stability of passive film with increase in potential. Transpassive zone or pitting zone represents the dissolution of passive film in the system. Based on the obtain PDP curve after 1 h of immersion, it was observed that steel in Sol. 2 shows only one zone with increase in potential i.e. This signifies that in contaminated environment due to low pH, solution would not be able to develop passivity on the rebar surface. On the other hand, a stable passive zone appeared up to 480 mV_{sce} in Sol. 1 specimen due to high pH of the solution. Also, potential of solution shifted to higher negative value that Sol. 1 represents higher corrosion activity on the surface of immersed specimen in Sol. 2 and formation of protective layer on specimen immersed in Sol. 1. From comparing Sol. 1 and Sol. 2, chloride in combination with low pH can causes rapid deterioration of the steel bar.

Figure 4 shows the polarization behaviour of specimens immersed in different solution tested after 240 h of long immersion. It can be seen from the Fig. 4 that polarization behaviour of curves similar for all tested solutions which represents absence of any

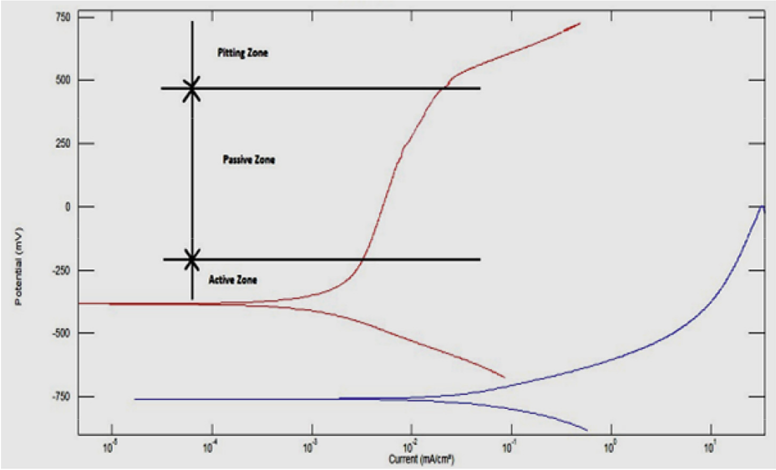


Fig. 3. Polarization curve of Sol. 1 and Sol. 2 after 24 h of immersion

stable film on the rebar surface except Sol. 4 i.e. 2-Aminopyridine that shows some passive region with change in potential. Also it can be seen that with addition of generic chemical polarization curves shifted to low i_{corr} values in comparison to Sol. 2. The value of i_{corr} for steel in different solution after 240 h along with inhibition efficiency of different generic compounds against corrosion are presented in Table 3. Form the Table 3, it can be said that specimen in Sol. 2 is highly corroded whereas addition of generic compounds reduces the corrosion activity in corrosive environment. Among the used compounds Sol. 3 and Sol. 4 are found to be most efficient. The order of efficiency are: 2AP > SA > PA > ABA.

5 Discussion

The protective capability of various generic compounds were tested in chloride contaminated low pH environment. Based on the test result it was found that 2-Aminopyridne was more effective than other compounds. The reason of high inhibition efficiency depends upon its molecular structure. 2AP has two nitrogen atoms close to each other. Location of these two atoms near to each give them advantage to share their lone pair of electrons with Fe^{2+} ions and forms a chelate ring. Whereas PA has similar pyridine ring along with adjacent COOH group that contain O with lone pair of electrons. Thus the mechanism of inhibition of these two molecules are quiet similar and this helps them to get good inhibition efficiency in long immersion test. Sac has benzene ring along with two functional group attached at adjacent position on its molecular structure, whereas AbA acid has benzene ring attached with same number of functional groups. In spite of having similarity in structure there is significant difference in inhibition efficiency was observed. The reason for this difference can be attribute to the positioning of functional groups. In Sac, COOH and OH group are close to each other and this will help the compound to interact with Fe to form chelate, whereas in AbA, NH_2 and COOH positioned opposite to each other thus only one could be able to interact with metal at a time.

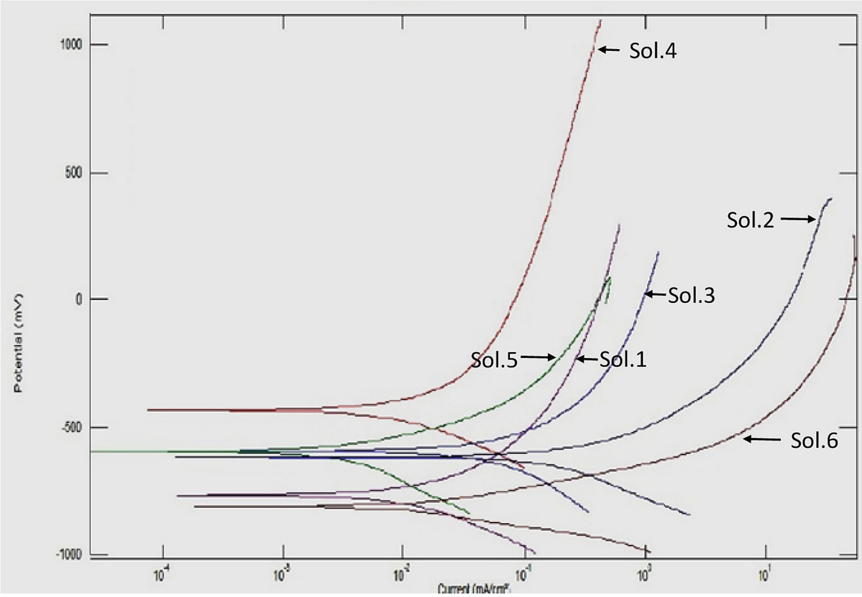


Fig. 4. Polarization curve of different solutions after 240 h of immersion

Table 3. Corrosion current density of different solution after 240 h and inhibition efficiency

Solutions	Sol. 1	Sol. 2	Sol. 3	Sol. 4	Sol. 5	Sol. 6
i_{corr} (mA/cm^2)	0.02	2.2	1.04	0.66	0.77	1.18
Efficiency (%)	–	–	53.7	70.5	66.0	47.8

6 Conclusion

On the basis of result obtain following conclusion can be drawn:

a. Presence of chloride in low pH environment can create more vulnerable situation for rebar and lead to severe corrosion.
b. Generic compounds can be an effective solution on retarding the corrosion activity in long exposure conditions. The reduction in i_{corr} suggest that all the inhibitors able to reduce the corrosion mechanism.
c. The inhibition efficiency was found at an order of AP > PA > Sac > AbA.
d. Inhibition of CI largely depends on the positioning of functional groups. Positioning of functional groups adjacent to each make compound more efficient as it provides opportunity for both the functional group to intract with metal ions.
e. Presence of amine and carboxylic group in compound can act as chelating agent and forms a barrier against the external corrosive agents.

References

1. Bowman, E., Jacobson, G., Koch, G., Varney, J., Thopson, N., Moghissi, O., Gould, M., Payer, J.: International measures of prevention, application, and economics of corrosion technologies study, p. A-19. NACE International (2016)
2. Aguirre-Guerrero, A.M., Mejía-De-Gutiérrez, R., Montês-Correia, M.J.R.: Corrosion performance of blended concretes exposed to different aggressive environments. Constr. Build. Mater. **121**, 704–716 (2016). https://doi.org/10.1016/j.conbuildmat.2016.06.038
3. Monticelli, C., Frignani, A., Balbo, A., Zucchi, F.: Influence of two specific inhibitors on steel corrosion in a synthetic solution simulating a carbonated concrete with chlorides. Mater. Corros. **62**, 178–186 (2011). https://doi.org/10.1002/maco.201005764
4. Ye, H., Jin, X., Fu, C., Jin, N., Xu, Y., Huang, T.: Chloride penetration in concrete exposed to cyclic drying-wetting and carbonation. Constr. Build. Mater. **112**, 457–463 (2016). https://doi.org/10.1016/j.conbuildmat.2016.02.194
5. Lee, M.K., Jung, S.H., Oh, B.H.: Effects of carbonation on chloride penetration in concrete. ACI Mater. J. **111**, 559–566 (2014). https://doi.org/10.14359/51685907
6. Guleria, H., Purnima, Tiwari, A.K., Goyal, S.: Performance of organic and inorganic functional groups as corrosion inhibitors in concrete experiencing extreme corrosive environment. Indian Concr. J. **95**, 29–37 (2021)
7. Tiwari, A., Goyal, S., Luxami, V., Chakraborty, M.K., Gundlapalli, P.: Evaluation of inhibition efficiency of generic compounds with additional heteroatom in simulated concrete pore solution and migration potential in concrete. J. Build. Eng. **43**, 102490 (2021). https://doi.org/10.1016/j.jobe.2021.102490
8. Tiwari, A.K., Purnima, S., Goyal, V.: Luxami, Influence of corrosion inhibitors on two different concrete systems under combined chloride and carbonated environment. Structures **48**, 717–735 (2023). https://doi.org/10.1016/j.istruc.2022.12.093
9. Das, J.K., Padhan, B.: Experimental investigation on inhibiting compounds against steel corrosion in concrete admixed with different chloride salts. Mater. Struct. **56**, 14 (2023). https://doi.org/10.1617/s11527-023-02103-1
10. Berdimurodov, E., Kholikov, A., Akbarov, K., Guo, L., Kaya, S., Katin, K.P., Verma, D.K., Rbaa, M., Dagdag, O., Haldhar, R.: Novel gossypol–indole modification as a green corrosion inhibitor for low–carbon steel in aggressive alkaline–saline solution. Colloids Surf. A Physicochem. Eng. Asp. **637**, 128207 (2022). https://doi.org/10.1016/j.colsurfa.2021.128207
11. El-Lateef, H.M.A., Khalaf, M.M., Gouda, M., Shalabi, K., El-Taib Heakal, F., Al-Janabi, A.S.M., Shaaban, S.: Novel water-soluble organoselenocyanates and symmetrical diselenides tethered *N*-succinanilate and *N*-maleanilate as corrosion inhibitors for reinforced steel in the simulated concrete pore solution. Constr. Build. Mater. **366**, 130135 (2023). https://doi.org/10.1016/j.conbuildmat.2022.130135
12. Kurtay, M., Gerengi, H., Chidiebere, A., Yıldız, M.: Corrosion inhibition of reinforcement steel in mixture water by caffeine and L-arginine. J. Adhes. Sci. Technol. **36**, 134–167 (2022). https://doi.org/10.1080/01694243.2021.1916251
13. El Ibrahimi, B., Jmiai, A., Bazzi, L., El Issami, S.: Amino acids and their derivatives as corrosion inhibitors for metals and alloys. Arab. J. Chem. **13**, 740–771 (2020). https://doi.org/10.1016/j.arabjc.2017.07.013
14. Aslam, R., Mobin, M., Zehra, S., Aslam, J.: A comprehensive review of corrosion inhibitors employed to mitigate stainless steel corrosion in different environments. J. Mol. Liq. **364**, 119992 (2022). https://doi.org/10.1016/j.molliq.2022.119992
15. Xu, Q., Hou, D., Zhang, H., Wang, P., Wang, M., Wu, D., Liu, C., Ding, Z., Zhang, M., Xin, Z., Fu, B., Guan, J., Zhang, Y.: Understanding the effect of vitamin B3 , B6 and C as a corrosion inhibitor on the ordinary Portland cement hydration: experiments and DFT study. Constr. Build. Mater. 331, 127294 (2022). https://doi.org/10.1016/j.conbuildmat.2022.127294

16. Hejazian, S., Ahmadyari-Sharamin, M.: Synergistic effect of Arabic and guar gums on corrosion behavior of steel reinforced concrete in a simulator environment. Case Stud. Constr. Mater. **16**, e01024 (2022). https://doi.org/10.1016/j.cscm.2022.e01024
17. Yang, H.-M.: Role of organic and eco-friendly inhibitors on the corrosion mitigation of steel in acidic environments—a state-of-art review. Molecules **26**, 3473 (2021). https://doi.org/10.3390/molecules26113473
18. Goyal, A., Ganjian, E., Pouya, H.S., Tyrer, M.: Inhibitor efficiency of migratory corrosion inhibitors to reduce corrosion in reinforced concrete exposed to high chloride environment. Constr. Build. Mater. **303**, 124461 (2021). https://doi.org/10.1016/j.conbuildmat.2021.124461
19. Osial, M., Wiliński, D.: Organic substances as corrosion inhibitors for steel in concrete—an overview, J. Build. Chem. **1**(1), 42–53 (2016). https://doi.org/10.17461/j.buildchem.2016.107
20. Ormellese, M., Lazzari, L., Goidanich, S., Fumagalli, G., Brenna, A.: A study of organic substances as inhibitors for chloride-induced corrosion in concrete. Corros. Sci. **51**, 2959–2968 (2009). https://doi.org/10.1016/j.corsci.2009.08.018
21. Tritthart, J.: Transport of a surface-applied corrosion inhibitor in cement paste and concrete. Cem. Concr. Res. **33**, 829–834 (2003). https://doi.org/10.1016/S0008-8846(02)01067-0
22. Söylev, T.A., Richardson, M.G.: Corrosion inhibitors for steel in concrete: state-of-the-art report. Constr. Build. Mater. **22**, 609–622 (2008). https://doi.org/10.1016/j.conbuildmat.2006.10.013
23. Gaidis, J.M.: Chemistry of corrosion inhibitors. Cem. Concr. Compos. **26**, 181–189 (2004). https://doi.org/10.1016/S0958-9465(03)00037-4
24. Ormellese, M., Bolzoni, F., Lazzari, L., Brenna, A., Pedeferri, M.: Organic substances as inhibitors for chloride-induced corrosion in reinforced concrete. Mater. Corros. **62**, 170–177 (2011). https://doi.org/10.1002/maco.201005763

Multi-criteria Decision-Making in Education—A Framework of Using Fuzzy Analytic Hierarchy Process for Study Program Selection

Dung Hai Dinh[1](✉), Ngoc Hong Tran[2](✉), and Vien Thuc Ha[3]

[1] Business Information Systems, Vietnamese-German University, Binh Duong, Vietnam
dung.dh@vgu.edu.vn

[2] Computer Science, Vietnamese-German University, Binh Duong, Vietnam
ngoc.th@vgu.edu.vn

[3] Faculty of Economics and Management, Vietnamese-German University, Binh Duong, Vietnam

Abstract. The process of selecting an individually appropriate undergraduate study program on university level has recently been essential for high school students, as the direct and indirect costs of studying, including opportunity costs are not trivial, and most students must spend a considerable investment of their lifetime and effort on the study chosen, for a duration over a few years. High school students and also universities can save a lot of money and time if they can identify the suitable choice of study program for each specific case and gain a competitive advantage by choosing the most suitable program that matches the profile of each specific student. Furthermore, as business environment and the required skill set is getting more complicated, there are also several different study programs available on the market. With multiple competing criteria, and erroneous parameters, our decision-making situation becomes more difficult, uncertain and ambiguous. As a result, a popular multiple-criteria decision-making technique, namely Fuzzy Analytic Hierarchy Process (AHP), can be employed as a solution to the issue of selecting study programs. This paper describes how Fuzzy AHP can be used to determine the best suitable study program based on the criteria that were chosen by a student. Our contribution not only applies the Fuzzy AHP to the study program selection problem, but also provides a comprehensive framework with the steps of Fuzzy AHP that can be used as a template for practical use.

Keywords: Study program selection · Multi criteria decision making · Fuzzy AHP · Education management · New Model University

1 Introduction

1.1 Background of the Problem

Education Management is about the coordination of activities that involve the administration and management of university, providing students with necessary knowledge and skills, and then preparing them for taking suitable job positions in the labor market.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 279–288, 2024.
https://doi.org/10.1007/978-3-031-50151-7_27

Thus, this kind of management includes the interrelationship between the university on one side as the provider of study programs and the student on the other side, who can be considered the customer that will "buy" the study program, the content of the curriculum and associated values of it.

In this paper we consider the concrete case of a new model University in Vietnam, which has intensive cooperation with a number of partner universities in Germany. Study programs which are offered at the University follow the Double Degree concept, meaning that students will receive two degrees, one from the Vietnamese university and one from a German partner university. At the time of enrollment, students will join the Foundation Year for one year and then have the chance to move on to the specified study program of their choice after that. Making this decision is based on several preferences and criteria, representing a typical Multi Criteria Decision Making problem in Education Management. Every year starting from February, a relatively large number of admission events are organized in Vietnam, offering consultation for high school students in their decision regarding which universities and study programs they should choose for their career. Thus, the framework presented in this work can contribute to the consultation and admission process at universities.

1.2 Problem Definition

One of the key issues of education management is the study program selection problem, because not only the expenditure during the study time but also the study time itself constitutes the major cost of a person lifetime. Study program selection is a multi-criteria challenge with qualitative and quantitative components. To choose the best suitable study program, a trade-off between these tangible and intangible components is required, some of which contain complicated, contradictory information that often reflect diverse points of view, and frequently change over time. Most situations are multi-criteria in nature, which means there are numerous evaluation criteria, preferences, and decision options. Information gathering can provide insight into the quantity and type of criteria that will be needed for review, as well as the type of data accessible. Making observations without specified criteria and measures may result in wasted work.

Without loss of generality, we deal with an example using 5 criteria and 3 study programs. The concrete criteria used in our model are chosen based on higher education expertise and listed below. Notice that the list of criteria can be easily adjusted to meet the realistic need of the high school students and their parents as well as ongoing changes in the academic environment.

1. Reputation: The reputation the granted degree of the German partner university.
2. Exchange: The possibility of having one exchange semester abroad in Germany.
3. Scholarship: The probability of being to receive a scholarship opportunity.
4. Attractiveness: The perceived attractiveness of the academic curriculum.
5. Employability: The possibility of getting a well-paid job shortly after graduation.

The assessment and scores of these criteria are different from one study program to another in nature. After defining those primary attributes for the selection problem, the weights and ratios are calculated using Fuzzy Analytic Hierarchy Process (AHP) approach. As a result, the study program with the highest priority weight will be chosen as the best suitable alternative. Thus, in this paper we set three objectives of research:

1. Model the study program selection problem as a multi-criteria decision-making problem.
2. Provide a template for a framework using Fuzzy AHP for solving the problem.
3. Illustrate the solution process with a concrete and understandable numerical example that can inspire further extension of the research. The final decision as an optimal solution is comprehensive and interpretable for students.

2 Literature Review

A popular decision-making method in Operations Research is the analytic hierarchy process (AHP). The decision is made in a final step after having derived the weights of the evaluating criteria (Saaty 1980). A lot of applications have been introduced, among which the supplier selection problem was popular, see Ghodsypour and O'Brien (1998), Wang et al. (2009), Yahya and Kingsman (1999), or Kilic (2013). The problem is that the AHP has been used for importance weighting. In addition, TOPSIS, a multi-criteria decision-making tool, can be a choice as described in Hwang and Yoon (1981), or DEA, which is a performance analysis tool, see Dinh and Le (2019). However, for the vagueness and uncertainty in the importance attributed to judgement of customer requirements, the crisp pairwise comparison in the conventional AHP is inadequate and incorrect to achieve the importance of customer requirements, see Chen et al. (2006), Cheng (1997), Ruoning and Xiaoyan (1992).

A broad overview of the fuzzy AHP methods existing in the literature and different approaches of fuzzy AHP methods with illustrating numerical examples can be found in the work of Demirel et al. (2008). Researchers in many different domains have employed Fuzzy AHP in project selection. They used weights to assign to the selected project qualities or criteria depending on their relevance. Kahraman et al. (2004) applied Fuzzy AHP to select the best catering firm to get customers the most satisfied. They interviewed customers of three Turkish catering firms and customers' considered the most important. Then a model is designed to select the best catering firm. Bilgen and Sen (2012) created a selection method for six sigma projects using a Fuzzy AHP. The primary characteristics of their Fuzzy AHP project selection technique were resources, benefits, and effects. Perez-Lespier et al. (2019) applied Fuzzy AHP in ranking the impact parameters to the environment in maritime transportation systems in order to assist decision makers to improve the environmentally sustainable maritime shipping. In education sector, Savić et al. (2013) was the first to mention the recommendation of appropriate study programs, based on the assessment of selected subjects.

3 The Fuzzy AHP Methodology

One of the first Fuzzy AHP applications was investigated by Van Laarhoven and Pedrycz (1983). In their work, the triangular membership functions for pair-wise comparisons are defined. Then, based on their work, Buckley (1985) calculated the fuzzy precedence of comparison ratios using triangular algorithms. For both of the criterion and the alternatives, Buckley's methods were employed to establish the relevance weights. Although J. Buckley's defuzzification approach was introduced many years ago, it is still valued for its efficiency and accuracy. The details of Fuzzy AHP are explained as follows.

Step 1: Create the hierarchical structure as shown in the following Fig. 1.

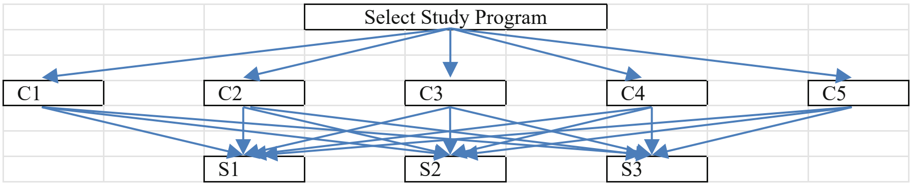


Fig. 1. Illustration of the analytic hierarchy process

Step 2: Construct the Fuzzy pairwise comparison matrix.

For transforming the language variable into a Fuzzy number, the transformation scale is used. The transition ranges for this study's scale of 1 to 9, see Fig. 2.

Definition	Fuzzy Triangular Scale
Equally important (Eq. Imp.)	(1, 1, 1)
Weekly important (W. Imp.)	(2, 3, 4)
Fairly important (F. Imp.)	(4, 5, 6)
Strongly important (S. Imp.)	(6, 7, 8)
Ansolutely important (A. Imp.)	(9, 9, 9)

Fig. 2. The transition ranges of the FAHP transformation scale

For each criterion $\tilde{d}_{\mathrm{ij}}$, we will use the average approach to compute the average score from the corresponding triangular fuzzy numbers. Assume there are k experts who estimate the importance of the criteria.

$$\tilde{d}_{\mathrm{ij}} = \frac{\sum_{k=1}^{K} \tilde{d}_{\mathrm{ij}}^{k}}{K} \tag{1}$$

From there, the Fuzzy pair-wise comparison matrix is built as follows.

$$\tilde{A}^{k} = \begin{bmatrix} \tilde{d}_{11}^{k} & \tilde{d}_{12}^{k} & \ldots & \tilde{d}_{1n}^{k} \\ \tilde{d}_{21}^{k} & \ldots & \ldots & \tilde{d}_{2n}^{k} \\ \ldots & \ldots & \ldots & \ldots \\ \tilde{d}_{n1}^{k} & \tilde{d}_{n2}^{k} & \ldots & \tilde{d}_{nn}^{k} \end{bmatrix} \tag{2}$$

Step 3: Calculate the Fuzzy weights.

The next step is to determine the fuzzy geometric mean and fuzzy weight for each criterion using the geometric mean approach. The geometric importance of each criterion's fuzzy comparison values is calculated as follows.

$$\tilde{r}_j = \left(\prod_{j=1}^{n} \tilde{d}_{\mathrm{ij}}\right)^{1/n}, \quad i = 1, 2, \ldots, n \tag{3}$$

Thus

$$\tilde{r}_j = (\tilde{a}_{i1} \times \tilde{a}_{i2} \times \tilde{a}_{i3} \times \ldots \times \tilde{a}_{in})^{1/n} \tag{4}$$

Step 4: Find the vector summation of each triangular values. Then, for each criterion I, we inverse the vector summation to achieve the fuzzy weight denoted as $\tilde{w}_{\mathrm{j}}$, then, we get the result multiplied with each triangular values to output the fuzzy weights below.

$$\begin{aligned} \tilde{w}_j &= \tilde{r}_j \times (\tilde{r}_1 + \tilde{r}_2 + \cdots + \tilde{r}_i + \cdots \tilde{r}_n)^{-1} \\ &= \left(L_{w_j}, M_{w_j}, U_{w_j}\right) \end{aligned} \tag{5}$$

With $\tilde{w}_j = \left(L_{w_j}, M_{w_j}, U_{w_j}\right)$, $L_{w_j}, M_{w_j}, U_{w_j}$ representing the lowest, average and highest values of the Fuzzy weights of the j^{th} criterion.

Step 5: Fuzzy weighted defuzzification.

$\tilde{w}_j$ are fuzzy numbers. Therefore, we de-fuzzified them by applying the mean of area method:

$$\overline{W_j} = \frac{L_{w_j}, M_{w_j}, U_{w_j}}{3} \tag{6}$$

With $\overline{W_j}$ being the real weight of j^{th}.

Step 6: $\overline{W_j}$ is a non-fuzzy number.The conversion of $\overline{W_j}$ to the weighted form is calculated by the following formula.

$$w_j = \frac{\overline{w_j}}{\sum_{i=1}^{n} \overline{w_j}} \tag{7}$$

With n being the total number of criteria.

As for the final step, the total scores for each alternative are computed by multiplying the weight of each alternative by the corresponding criteria. According to the results, the option with the greatest score is chosen. A case study about study program selection for students is provided in the next part to clarify the methodology and demonstrate its usefulness.

4 Application and Interpretation of Results

For an example using 5 criteria and 3 study programs, as illustrated in Fig. 3, values of the same criterion might be equal or different from one study program to another, resulting in a relative comparison matrix that will be built in this section.

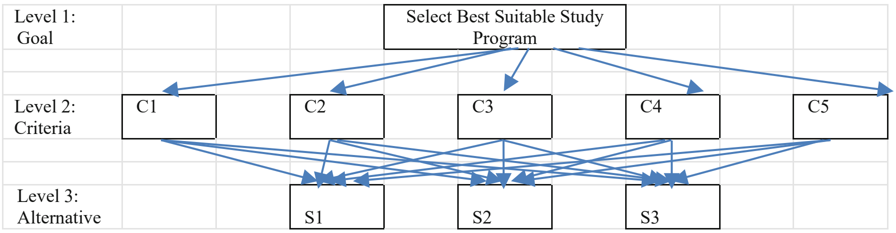


Fig. 3. The hierarchical structure of the selection problem

Table 1. Pair-wise comparisons versus selected criteria

A. Imp (9,9,9)	S. Imp (6,7,8)	F. imp (4,5,6)	W. Imp (2,3,4)	Criteria	Eq. Imp (1,1,1)	Criteria	W. Imp (2,3,4)	F. imp (4,5,6)	S. Imp (6,7,8)	A. imp (9,9,9)
			X	Reputation		Exchange				
	X			Reputation		Scholarship				
		X		Reputation		Attractiveness				
			X	Reputation		Employability				
		X		Exchange		Scholarship				
				Exchange		Attractiveness	X			
			X	Exchange		Employability				
	X			Scholarship		Attractiveness				
			X	Scholarship		Employability				
				Attractiveness		Employability		X		

The header labels are described in Fig. 2.

The five criteria chosen for this problem, as mentioned in Sect. 1, are Reputation, Exchange, Scholarship, Attractiveness, and Employability. The three study programs can be taken from the Faculty of Engineering, such as Electrical Engineering, Industrial Engineering, and Mechanical Engineering. A quick survey and interview was done for the student and assessment result was collected from the consultation process with other educators and education managers to average the pair-wise comparison of the criteria, based on their preferences we decide the important of criteria compared to others as presented in Table 1. From this Table 1, we get the translation of the pairwise comparison in form of a comparison matrix for these criteria. The following Table 2 presents this matrix.

After computing the fuzzy weights at step 3 above, the geometric centre of each criterion's fuzzy comparison values, the total values and reverse values are calculated (see Table 3).

Table 2. The comparison matrix for criteria

Criteria	Reputation	Exchange	Scholarship	Attractiveness	Employability
Reputation	(1,1,1)	(2,3,4)	(6,7,8)	(4,5,6)	(2,3,4)
Exchange	(1/2,1/3,1/4)	(1,1,1)	(4,5,6)	(1/2,1/3,1/4)	(2,3,4)
Scholarship	(1/6,1/7,1/8)	(1/4,1/5,1/6)	(1,1,1)	(6,7,8)	(2,3,4)
Attractiveness	(1/4,1/5,1/6)	(2,3,4)	(1/6,1/7,1/8)	(1,1,1)	(4,5,6)
Employability	(1/2,1/3,1/4)	(1/2,1/3,1/4)	(1/2,1/3,1/4)	(1/4,1/5,1/6)	(1,1,1)

Table 3. Geometric importance of of each criterion's fuzzy comparison values

Criteria	$\tilde{r}_j$		
Reputation	2.491	3.160	3.776
Exchange	1.149	1.108	1.084
Scholarship	0.871	0.903	0.922
Attractiveness	0.803	0.844	0.871
Employability	0.500	0.375	0.304
Total	5.813	6.389	6.958
Reverse	0.172	0.157	0.144
Increasing order	0.144	0.157	0.172

Similarly, after calculating the geometric mean of all criterion's values, the fuzzy weight of all criteria can be found. From there, we get the Averaged relative weights and Normalized relative weights of criteria, see Tables 4 and 5.

Table 4. Relative fuzzy weights

Criteria	$\tilde{w}_j$		
Reputation	0.358	0.495	0.650
Exchange	0.165	0.173	0.187
Scholarship	0.125	0.141	0.159
Attractiveness	0.115	0.132	0.150
Employability	0.072	0.059	0.052

Now the same process is applied to discover the relevant values for the study programs, after having achieved the relative normalized non-fuzzy weights for the criteria. The options, i.e. study programs, should now be assessed in pairs with respect to each

Table 5. Mean and normalized weights for each criteria

Criteria	$\overline{w}_j$	w_j
Reputation	0.501	0.495
Exchange	0.175	0.173
Scholarship	0.142	0.140
Attractiveness	0.132	0.131
Employability	0.061	0.060

criterion. The analysis is repeated 5 times for 5 criteria. Calculation results for the Reputation criterion are as shown as an example in the two following Tables 6 and 7. For the other four criteria, similar steps must be done.

Table 6. Reputation pair wise comparisons

A. Imp (9,9,9)	S. Imp (6,7,8)	F. Imp (4,5,6)	W. imp (2,3,4)	Repuatation	Eq. Imp (1,1,1)	Reputation	W. Imp (2,3,4,)	F. Imp (4,5,6,)	S. imp (6,7,8)	A. Imp (9,9,9)
			X	Study program 1		Study program 2				
				Study program 1		Study program 3			X	
				Study program 2		Study program 3		X		

The header labels are described in Fig. 2

Table 7. Averaged and normalized relative weights calculated for “reputation”

Reputation	$\overline{w}_j$	w_j
Study program 1	0.170	0.169
Study program 2	0.094	0.093
Study program 3	0.745	0.738

The summary of all calculated relative weights of five criteria are displayed in Table 8, whereas the final result of decision making is provided in Table 9.

Study program 3 has the highest total score based on this outcome. As a result, it is recommended as the best alternative out of three, based on five selected criteria and the decision makers’ fuzzy preferences which were defined at the beginning.

5 Conclusion and Future Works

We applied Fuzzy AHP method for choosing among a number of alternatives under conflicting decision criteria. Here the alternatives are the study programs that can be chosen by high school students. The Fuzzy AHP model used in this study helps provide

Table 8. Normalized non-fuzzy relative weights of all five criteria

Alternatives	Reputation	Exchange	Scholarship	Attractiveness	Employability
Study program 1	0.169	0.617	0.726	0.290	0.244
Study program 2	0.093	0.086	0.076	0.273	0.390
Study program 3	0.738	0.297	0.198	0.437	0.366

Table 9. Aggregated results for final decision

Criteria	Weights	Study program 1	Study program 2	Study program 3
Reputation	0.495	0.169	0.093	0.738
Exchange	0.173	0.617	0.086	0.297
Scholarship	0.140	0.726	0.076	0.198
Attractiveness	0.131	0.290	0.273	0.437
Employability	0.060	0.244	0.390	0.366
Total		0.345	0.131	0.524
Rank		2	3	1

a scientific approach to evaluate and discover potential and suitable study programs for students. The model allows an evaluation using a variety of criteria while limiting the subjectivity of decision makers. According to the ranking results, study program 3 has the greatest probability to be suitable for the student, followed by programs 1 and 2 respectively. With this result and the provided framework, the paper offers a good insight into one of the most common problems that arise repeatedly on a yearly basis, namely study program selection for high school students.

There are limitations and potentials for future studies. In education, study program criteria, individual strengths and weaknesses, and personal preferences are the most common types of selection criteria. The fact that the criteria for assessing competitive advantage are established on surveys and gathering opinions from educational experts makes it hard to assess the adequacy of the criteria. This could potentially form the basis for future studies. Some criteria may be difficult to assess throughout the selecting process, e.g. Availability of lab equipment, Supporting services, or Management and soft skills. Information can be difficult to generate, complicated to analyze, as well as time-consuming. An extension of the framework can be done to deal with more criteria and study programs, as well as considering big data analytics. Furthermore, there can be other Operations Research methods that can be used for either the purpose of prediction using Markov Model to test the future suitability of study programs, see Dung et al. (2022), or a combination of AHP or ANP and other mathematical programming techniques to optimize the final result based on the problem instance and preference of the decision maker.

References

Bilgen, B., Sen, M.: Project selection through fuzzy analytic hierarchy process and a case study on six sigma implementation in an automotive industry. Prod. Plan. Control **23**, 2–25 (2012)

Buckley, J.: Fuzzy hierarchical analysis. Fuzzy Sets Syst. **17**(1), 233–247 (1985)

Chen, C.T., Lin, C.T., Huang, S.F.: A fuzzy approach for supplier evaluation and selection in supply chain management. Int. J. Prod. Econ. **102**(2), 289–301 (2006)

Cheng, C.H.: Evaluating naval tactical missile system by fuzzy AHP based on the grade value of membership function. Eur. J. **96**(2), 343–350 (1997)

Demirel, T., Demirel, N.C., Kahraman, C.: Fuzzy analytic hierarchy process and its application: theory and applications with recent developments. In: Fuzzy Multi-Criteria Decision Making. Springer Optimization and Its Applications, vol. 16, pp. 53–83. Springer, Boston, MA (2008)

Dinh, H.D., Le, K.M.: On the use of data envelopment analysis to improve performance efficiency of governmental management in big cities. In: Huong, L.T.T., Pomeroy, G.M. (eds.) AUC 2019. Advances in 21st Century Human Settlements, pp. 149–157. Springer, Singapore (2021)

Dung, D.H., Quang, N.M., Chi, B.H.: A Markov chain model for predicting brand switching behavior toward online food delivery services. In: Thach, N.N., Kreinovich, V., Ha, D.T., Trung, N.D. (eds.) Financial Econometrics: Bayesian Analysis, Quantum Uncertainty, and Related Topics. Studies in Systems, Decision and Control, vol. 427, pp. 781–797. Springer, Cham (2022)

Ghodsypour, S.H., O'Brien, C.: A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming. Int. J. Prod. Econ. **56–57**(20), 199–212 (1998)

Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making Methods and Applications: A State of the Art Survey. Springer Verlag, USA (1981)

Kahraman, C., Cebeci, U., Ulukan, Z.: Multi-criteria supplier selection using fuzzy AHP. Logist. Inf. Manag. **16**(6), 382–394 (2003)

Kahraman, C., Cebeci, U., Ruan, D.: Multi-attribute comparison of catering service companies using fuzzy AHP: the case of Turkey. Int. J. Prod. Econ. **87**, 171–184 (2004)

Kilic, H.S.: An integrated approach for supplier selection in multi item/multi supplier environment. Appl Math Model. **37**(14–15), 7752–7763 (2013)

Kwong, C.K., Bai, H.: A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment. J. Intell. Manuf. **13**, 367–377 (2002)

Pérez Lespier, L., Long, S., Shoberg, T., Corns, S.: A model for the evaluation of environmental impact indicators for a sustainable maritime transportation systems. Front. Eng. Manage. **6**(3), 368–383 (2019). https://doi.org/10.1007/s42524-019-0004-9

Ruoning, X., Xiaoyan, Z.: Extensions of the analytic hierarchy process in fuzzy environment. Fuzzy Sets and Syst. **52**(3), 251–257 (1992)

Saaty, T.L.: The Analytic Hierarchy Process. McGraw Hill, New York (1980)

Savić, G., Makajić-Nikolić, D., Ranđelović, D., Ranđelović, M.: Study program selection by aggregated DEA-AHP measure. Metal. Int. **18**(1), 169–174 (2013)

Van Laarhoven, P.J.M., Pedrycz, W.: A fuzzy extension of saaty's priority theory. Fuzzy Sets Syst. **11**(1–3), 229–241 (1983)

Wang, J.W., Cheng, C.H., Cheng, H.K.: Fuzzy hierarchical TOPSIS for supplier selection. Appl. Soft Comput. **9**(1), 377–386 (2009)

Yahya, S., Kingsman, B.: Vendor rating for an entrepreneur development programme: a case study using the analytic hierarchy process method. J. Oper. Res. Soc. **50**, 916–930 (1999)

The Cost Optimization of Reinforced Concrete Retaining Wall with Different Algorithms

Muhammed Çoşut, Gebrail Bekdaş, and Sinan Melih Nigdeli(✉)

Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcılar, Istanbul, Turkey
{bekdas,melihnig}@iuc.edu.tr

Abstract. In this study, the cost optimization of the retaining wall under distributed load was carried out using Matlab program with different metaheuristic algorithms to provide the boundary conditions of the regulation. Various metaheuristic algorithms which are Jaya, TLBO (Teaching-Learning Based Optimization) and Hybrid (TLBO-Jaya) were used, which can be effective in solving more complex problems were compared. The use of such different types of algorithms provides a comparison between the algorithms such as approximately at which iteration it reaches the objective function and the value created in each step while reaching the objective function by calculating the standard deviation value. In this way, comparisons were made between the optimization types and which one was more efficient and approached the objective function in fewer iterations. In the optimization phase, the sections of the retaining wall were defined as variables in the system and in accordance with the regulations when these variables were entered at specified intervals. As a result of these studies, while it is seen that the Hybrid Algorithm reaches the objective function with the least iteration, it also has a small standard deviation value. It is found that the Hybrid algorithm is more effective in the process of retaining wall problems.

Keywords: Metaheuristic algorithm · TS500 · Optimization · Optimum cost design · Jaya · TLBO · Hybrid algorithm

1 Introduction

After the settled life, the fact that some conditions were provided for the accommodation of people emerged. Therefore, the use of many materials has created systems in different shapes, sizes and with different materials. Day by day, studies are carried out around the world that will provide significant progress in every field and contribute to the development of existing elements or systems. Civil engineering is also one of the areas that are greatly affected by this situation, before this development, the majority of designers adjust the cross-sections by their experiments [1, 2]. Development and change have a significant impact on construction systems and building materials, ensuring the formation of more efficient and sustainable systems. The basic principle of these systems is that, in addition to being more economical than existing building systems, it is expected

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 289–298, 2024.
https://doi.org/10.1007/978-3-031-50151-7_28

to be better depending on the situation, which will be considered as strength capacity, displacement and stability [3]. Considering these situations, academic studies are carried out and the effects of positive and negative results are detailed by observing the effects of many materials such as fly ash and carbon fiber used in building systems. Recently, with the development of some machines, these operating situations have started to increase. For some engineers, the diversity of computer programs and the availability of different useful features as well as analysis of different structural components are undertaken to look at performance levels and applicability. Retaining walls can be an example of this structural system, and they hinder the soil slopes which can be in railways, bridges [4] and also road constructions [5]. Reinforced concrete retaining walls consist of a body and a foundation plate which can be divided into heel and toe sections [6]. Kalemci et al. [7], the lowest weight cantilever retaining wall design in accordance with the regulations has been realized by using Gray Wolf optimization. Temür and Bekdaş [8] optimization of the cantilever retaining wall was carried out using particle swarm optimization, Harmony Search optimization and Learning-Teaching Based Optimization. In addition, in order to optimization, Ceranic and Fryer [9] used Simulated Annealing, Khajehzadeh et al. [10] used Particle Swarm Optimization.

Within the scope of this study, the retaining wall's design was optimized as a cost-objective function. The majority of researchers focus on estimating construction costs in the first stages of the construction design processes [11–17]. While performing this optimization, 3 different Metaheuristic Algorithms in Matlab program were run 10 times for each algorithm at a certain iteration value, and the resulting data were recorded. While the recorded data shows the average iteration value of the algorithms reaching the objective function, with graphics, the standard deviation values of the value that will occur in each cycle of these algorithms will be found and displayed in the form of graphs.

2 Methodology

2.1 The Design of Retaining Wall

Retaining walls are one of the building elements and can be designed in different sizes and shapes by using different materials. These various designs differ according to the seismicity of the location [18], the intended use of the system, the magnitude of the incoming loads and also certain size situations. In general, the commonly used retaining wall system is reinforced concrete retaining walls. This system can be applied according to different features and designs such as buttresses and heels. These designs can contribute to properties separately, and for example, some systems increase the interaction with the soil, some of them increase friction, some of them prevent slippage and also some designs work more efficiently in case of overturning. However, all retaining walls designed in general show an effective performance as a result of loadings and possible situations [19] by fully meeting the design criteria of the country in which they are made. With the use of many software and programs, the cross-sectional dimensions are reached by making analyzes by accepting every possible effect and the necessary impacts without applying such different kinds of designs.

Figure 1 illustrates the retaining wall control situations which are sliding as well as turning. There is various effect on the retaining wall as a load which influences the retaining wall in different ways as seen in Fig. 2. All retaining wall generally are constructed according to passive force, active force, surcharge load effect, concrete weight and also soil weight.

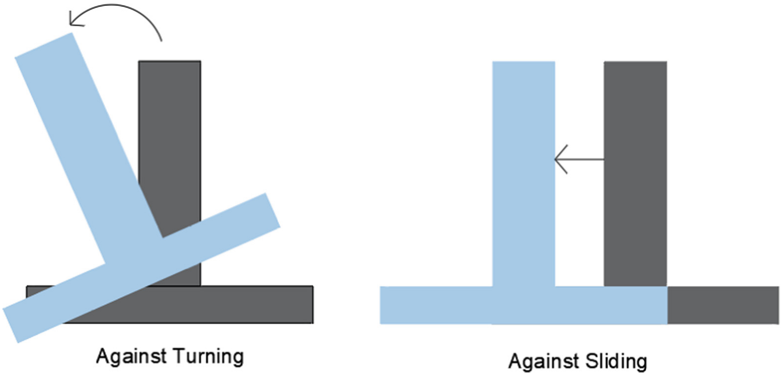


Fig. 1. Retaining wall control circumstances

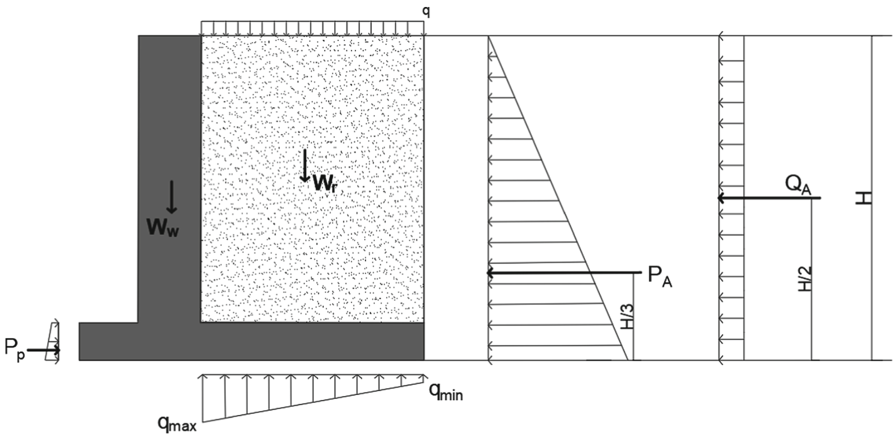


Fig. 2. Distribution of load on retaining wall

It is required to calculate the retaining wall against turning and against sliding [20] according to Eqs. (1) and (2), respectively. Equation (3) finds eccentric values. Equation (4) includes 2 equations to find the maximum and minimum force on retaining.

M_R is the sum of the moments that resist overturning, M_O *is* the sum of the moments that caused the overturning, N is the normal force, M is the bending moment, F_R forces that resist slipping as well as F_D is the force sliding the retaining wall.

L is the length the heel of retaining wall, and V is used for weight of concrete.

$$FS_{overturning} = \frac{\sum M_R}{\sum M_o} \tag{1}$$

$$FS_{sliding} = \frac{\sum F_R}{\sum F_D} \tag{2}$$

$$\mathrm{e} = \frac{L}{2} - x' \quad \text{and} \quad \mathrm{e} = \frac{L}{2} - \frac{\sum M_R - \sum M_o}{\sum V} \leq \frac{L}{6} \tag{3}$$

$$\mathrm{q}_{\mathrm{max,min}} = \frac{\sum V}{L} \times \left(1 \pm \frac{6e}{L}\right) \quad \text{or} \quad \mathrm{q}_{\mathrm{max,min}} = \frac{N}{L} - 6 \times \frac{M}{L^2} \tag{4}$$

K_p is defined as the coefficient of soil pressure in the passive state (Eq. (5)) while K_a is defined as the coefficient of soil pressure in the active state (Eq. (6)). Passive (Eq. (7)) and active force (Eq. (8)) are necessary to find the design of the retaining wall.

$$\mathrm{K_p} = \tan^2\left(45 - \frac{\text{Ø}'}{2}\right) \tag{5}$$

$$\mathrm{K_a} = \cos\beta \frac{\cos\beta - \sqrt{\cos^2\beta - \cos^2\theta}}{\cos\beta + \sqrt{\cos^2\beta - \cos^2\theta}} \tag{6}$$

$$\mathrm{P_a} = \frac{1}{2}\gamma H^2 K_a \tag{7}$$

$$\mathrm{P_p} = \frac{1}{2}\gamma H_P^2 K_p \tag{8}$$

2.2 The Used Algorithms with the Optimization Process

In this study, the Hybrid, Jaya algorithm as well as Teaching-Learning Based Optimization are used. These have different processes to reach objective functions which can be related to cost, CO_2 emission or system displacement and so on.

Jaya algorithm was found by Rao [21] in 2016. It does not have specific control parameters, and it is easy to use [22]. Jaya algorithm equation can be seen as Eq. (9). This equation consists of $X'_{i,new}$ (new value), $X_{i,g_{best}}$ (best value for initial design variables) and $X_{i,g_{worst}}$ (worst value for initial design variables).

$$X'_{i,new} = X_{i,j} + rand()\left(X_{i,g_{best}} - \left|X_{i,j}\right|\right) - rand()\left(X_{i,g_{worst}} - \left|X_{i,j}\right|\right) \tag{9}$$

Teaching-Learning Based Optimization (TLBO) was founded by Rao in 2011 [23]. This algorithm has a 2-stage control phase. Therefore, it can enable to process to reach a lower iteration. TF is the Teaching Factor which can be found easily in Eq. (10). When TF is found, new variable values can be found in Eq. (11) for the teaching phase. $X_{i,mean}$ is the mean of existing solutions.

$$\mathrm{TF} = round(1 + rand()) \tag{10}$$

$$X_{i,new} = X_{i,j} + rand()\left(X_{i,g_{best}} - \left|X_{i,j}\right|\right) - (TF)X_{i,mean}) \tag{11}$$

When consequently passing the learning phase, there are 2 different occasions which are about a and b, and these variables are generated randomly. If a is smaller than b, the first of Eq. (12) will be used. Otherwise, the second is used. After this phase, the iteration number is checked by code and if it completes the maximum iteration number, it will finish the optimization process.

$$X_{i,new} = \begin{cases} AF_a < AF_b, \; X_{i,j} + rand()\left(X_{i,a} - X_{i,b}\right) \\ AF_a > AF_b, \; X_{i,j} + rand()\left(X_{i,b} - X_{i,a}\right) \end{cases} \tag{12}$$

TLBO-Jaya is a combined hybrid algorithm [24]. Jaya algorithm was added in TLBO and the learning phase was removed.

3 Numerical Example

Table 1 consists of variables, constants and material cost. Generally, some values have different limitations such as clear cover (P_c) should be a minimum of 40 mm. These values are taken as given in Table 1.

Table 1. Design constraints, variables and material cost

Explanation	Symbol	Unite	Values
Maximum base plate thickness	h_{wmin}	mm	300
Minimum base plate thickness	h_{wmak}	mm	600
Maximum console breadth	h_{bmin}	mm	300
Minimum console breadth	h_{bmak}	mm	600
Maximum length of the heel	h_{min}	mm	1500
Minimum length of the heel	h_{maks}	mm	4000
Unit soil weight	γ_s	kN/m^3	18
Surcharge load	q	kN/m	10
Soil bearing capacity behind the wall	q_u	kPa	300
Soil internal friction angle	$\O_R$	°	30
Compressive strength of concrete	fck	MPa	25
Yield strength of concrete	fyk	MPa	420
Specific density of steel	γs	t/m^3	7.86
Clear cover	Pc	mm	50
Concrete cost per unit volume	Cc	TL/m^3	790
Steel cost per unit weight	Cs	TL/ton	14,900

All variables are taken from the standard [25] and also the material cost given is the most up-to-date price. Variables interval are given as maximum-minimum length and materials property is given also. The nomenclature for retaining wall sizing is shown in Fig. 3 as h_b, h_w, L1, L2 and H. Moreover, L is equal to L1 + L2 + h_w.

$$Total\ Concrete\ Cost = C_c * Total\ volume\ of\ Concrete \tag{13}$$

$$Total\ Steel\ Cost = C_s * Total\ area\ of\ Steel \tag{14}$$

$$Total\ Cost = Total\ Concrete\ Cost + Total\ Steel\ Cost \tag{15}$$

Equations (13) and (14) show the total cost of concrete and steel respectively, and Eq. (15) illustrates the total cost for the design of the retaining wall.

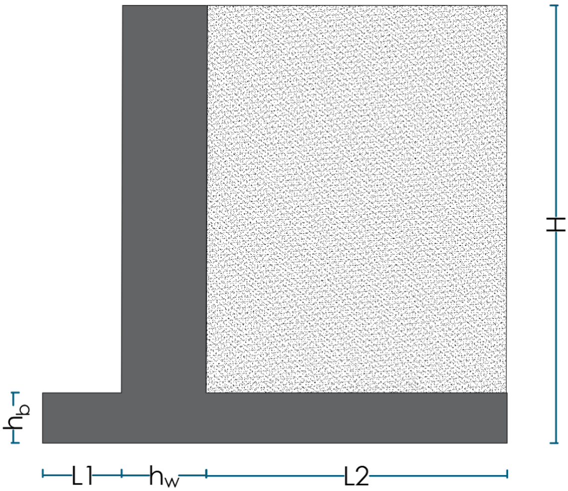


Fig. 3. Retaining wall cross section

The result of retaining wall optimization can be seen in Table 2. Looking at the table in more detail, the base plate thickness is found 0.313 m, while the console breadth is found as 0.317 m so they are similar values compared to each other. The rear encasement length is found as 3.21 m and also the length of the base is 3.53 m. Furthermore, the reinforcement area is 6300 mm^2 which is necessary to design constraints. The cost of this design is 5691 TL.

Table 3 shows the limitation results and it can be seen that all result of optimization design provides all constraints. Safety constraint values are needed to be smaller than optimization results. Furthermore, the reinforcement area should provide maximum and minimum intervals. Optimization results for resistance capabilities should be under the constraint value.

Every algorithm which is used in this study was run ten times. Therefore, an average of this 10 times running is found. Figure 4 compares the algorithms about how many iterations are needed to reach the objective function.

Table 2. The result of retaining wall optimization

Symbol	Unit	Results
hb	m	0.313
h_w	m	0.317
L2	m	3.21
L	m	3.53
A_s	mm^2	6300
F(x)	TL	5691

Table 3. Optimization results

Explanation	Limit values	Optimization results	
Safety for overturning	2	$\leq$	2.71
Safety for sliding	2	$\leq$	2.028
Safety of bearing capacity	1	$\leq$	1.13
q_u (the flexural resistance capabilities of critical sections)	300	$\geq$	265.75
qmin (minimum bearing stress)	0	$\leq$	5.7×10^{-14}
Minimum reinforcement areas	0.01	$\geq$	0.01
Maximum reinforcement areas	0.04	$\leq$	0.01

In Fig. 5, these 3 algorithms are shown as a comparison due to recognize which one is better than the other when they run and reach an objective function.

Table 4 demonstrates the cross-sections, for the exact objective function iteration number and standard deviation as well as the cost for Jaya algorithm, TLBO and TLBO-Jaya. The cross-sectional dimensions were found to be the same in three different algorithm types in accordance with the regulation conditions for retaining wall design. The same cross-sectional size and the same reinforcement area to be used ensured that the total cost values were also the same. It should be noted here that it is the controls of how the algorithms that reach the objective function converge to this value and how many iterations on average these results are achieved. When the values were checked, it was observed that the Hybrid algorithms reached the objective function in the fewest iterations.

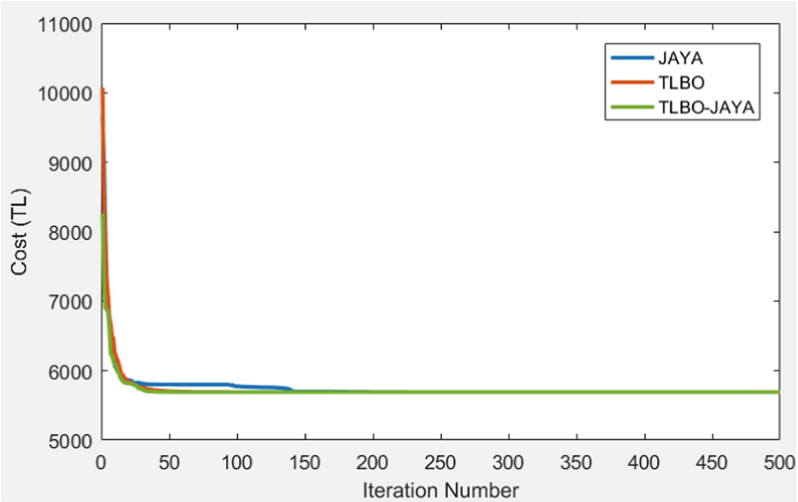


Fig. 4. Comparison between used algorithms for cost

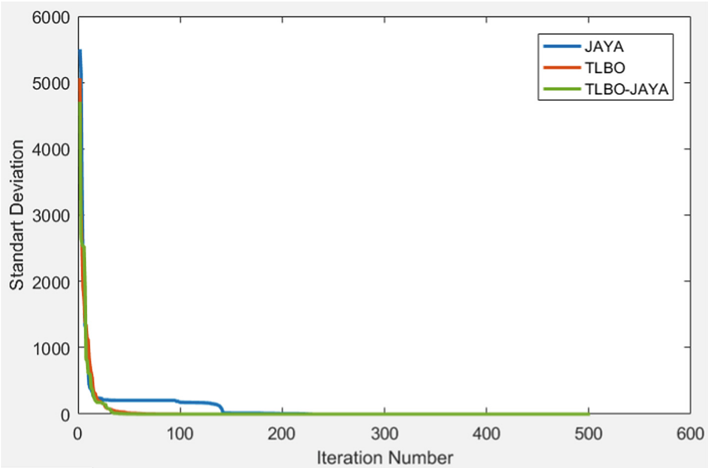


Fig. 5. Comparison between used algorithms for standard deviation

Table 4. Mean of 10 iterations for 3 algorithms

Explanation	JAYA	TLBO	Hybrid
hb (m)	0.317	0.317	0.317
h_W (m)	0.313	0.313	0.313
L_2 (m)	3.21	3.21	3.21
L (m)	3.53	3.53	3.53
Iteration number	268	414	215
Mean of standard deviation	99.5	52.2	48.27
Cost (TL)	5691	5691	5691

4 Conclusion

Retaining walls can be optimally designed in various ways and with various algorithms. Nevertheless, the differences between these algorithms just can be time which is about which iteration number and also approaching ratio to reach an objective function. Using 3 different algorithms contributes to comparing which algorithm is more usable and reaches the optimum result with the least iteration number. It can easily be seen that all 3 algorithms reached 5691 TL even though some groups result differently compared to each other. To exemplify this, iteration numbers that reach approximately the objective function are 268 iterations, 414 iterations and 215 iterations for Jaya, TLBO and TLBO-Jaya, respectively. The hybrid algorithm reached the objective function faster than other algorithms, followed by Jaya and TLBO. Although Jaya algorithm reached the objective function faster than TLBO, the mean of the standard deviation of TLBO which is 52.2 is better than Jaya algorithm. This means that TLBO converges to an objective function before and the numbers are usually approximately close to an objective function. Moreover, the Hybrid algorithm's standard deviation mean is 48.27 so it is best to use it for this problem. Generally, the hybrid algorithm and TLBO are better than Jaya algorithm because of the 2 phases they have. As a consequence, Hybrid algorithms can give much more efficient results in solving more complex problems, and better results will be obtained with the development of these algorithms.

References

1. Yepes, V., Alcala, J., Perea, C., Gonzalez-Vidosa, F.: A parametric study of optimum earth-retaining walls by simulated annealing. Eng. Struct. **30**, 821–830 (2008)
2. Kaveh, A., Behnam, A.: Charged system search algorithm for the optimum cost design of reinforced concrete cantilever retaining walls. Arab. J. Sci. Eng. **38**, 563–570 (2013)
3. Temür, R., Bekdaş, G.: Teaching learning-based optimization for design of cantiliever retaining walls. Struct. Eng. Mech. **57**(4), 763–783 (2016)
4. Gandomi, A.H., Kashani, A.R., Roke, D.A., Mousavi, M.: Optimization of retaining wall design using recent swarm intelligence techniques. Eng. Struct. **103**, 72–84 (2015)
5. Dey, A., Basudhar, P., Deb, K., Vashistka, A.: Cost optimization of reinforced earth walls. Geotech. Geol. Eng. **26**, 1–12 (2008)

6. Öztürk, H.T.: Betonarme Konsol İstinat Duvarların Minimum Maliyet ve Ağırlıkla Optimum Tasarımı. Düzce Üniversitesi Bilim ve Teknoloji Dergisi **6**, 1258–1268 (2018)
7. Kalemci, E.N., İkizler, S.B., Dede, T., Angın, Z.: Design of reinforced concrete cantilever retaining wall using grey wolf optimization algorithm. Structures **23**, 245–253 (2020)
8. Temür, R., Bekdaş, G.: Optimum design of reinforced concrete retaining walls. In: Handbook of Research on Predictive Modeling and Optimization Methods in Science and Engineering, p. 19 (2018)
9. Ceranic, B., Fryer, C.: An application of simulated annealing to optimum design of reinforced concrete retaining structures. J. Comput. Struct. **79**, 1569–1581 (2001)
10. Khajehzadeh, M., Taha, M.R., El-Shafie, A., Eslami, M.: Modified particle swarm optimization for optimum design of spread footing and retaining wall. J. Zhejiang Univ. Sci. A **12**(6), 415–427 (2011). https://doi.org/10.1631/jzus.A1000252
11. Dağdeviren, U., Kaymak, U.: Cost estimation models for the reinforced concrete retaining walls. BSEU J. Sci. **2458–7575**, 9–26 (2020)
12. Günaydın, H., Doğan, S.: A neural network approach for early cost estimation of structural systems of buildings. Int. J. Project Manag. **22**(7), 595–602 (2004)
13. Wang, X.Z., Duan, X.C., Liu, J.Y.: Application of neural network in the cost estimation of highway engineering. J. Comput. **5**(11), 1762–1766 (2010)
14. Arafa, M., Alqedra, M.: Early stage cost estimation of buildings construction projects using artificial neural networks. J. Artif. Intell. **4**(1), 63–75 (2011)
15. Chandanshive, V., Kambekar, A.: Prediction of early stage construction cost of building projects using artificial neural network. Int. J. Sci. Eng. Res. **5**(7), 453–463 (2014)
16. Gencer, H.: Estimated cost calculation guidelines, problems in practice and a solution recommendation. In: Uluslararası Katılımlı 7. İnşaat Yönetimi Kongresi-Samsun, pp. 387–398 (2017)
17. Juszczyk, M., Lesniak, A., Zima, K.: ANN based approach for estimation of construction costs of sports fields. Complexity 1–11, Article ID 7952434 (2018)
18. Choudhury, D., Nimbalkar, S.S.: Pseudo-dynamic approach of seismic active earth pressure behind retaining wall. Geotech. Geol. Eng. **24**, 1103–1113 (2006)
19. Qin, C., Chian, S.C.: Seismic Stability of Geosynthetic-reinforced walls with variable excitation and soil properties: a discretization-based kinematic analysis. Comput. Geotech. **102**, 196–205 (2018)
20. Kayabekir, A.E., Arama, Z.A., Bekdaş, G., Niğdeli, S.M., Geem, Z.W.: Eco-friendly design of reinforced concrete retaining walls: multi-objective optimization with harmony search applications. Sustainibility **12**, 6087 (2020)
21. Rao, R.V.: Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int. J. Ind. Eng. Comput. **7**, 19–34 (2016)
22. Bekdaş, G., Nigdeli, M.N., Yücel, M., Kayabekir, A.E.: Yapay Zeka optimizasyon Algoritmaları ve Mühendislik Uygulamaları, Seçkin Yayıncılık, Ankara, ISBN:978-975-02-6637-9 (2021)
23. Rao, R.V.: Teaching-learning based optimization: a novel method for constrained mechanical design optimization problems. Comput. Aided Des. **43**, 303–315 (2011)
24. Ting, T.O., Yang, X.-S., Cheng, S., Huang, K.: Hybrid metaheuristic algorithms: past, present and future. In: Recent Advances in Swarm Intelligence and Evolutionary Computation, Studies in Computational Intelligence, vol. 585: pp. 71–83 (2015)
25. TS500.: Betonarme Yapıların Tasarım ve Yapım Kuralları. Türk Standartları Enstitüsü, Ankara (2000)

Energy Efficient Distributed Cloud Load Balancing Using Virtual Machines

Sabbir Mahmud Afridi[1], Akmal Hossain[1], Md Mehdi Ahsan Rafsan[1], Ahmed Wasif Reza[1](✉), and Mohammad Shamsul Arefin[2](✉)

[1] Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
wasif@ewubd.edu

[2] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh
sarefin@cuet.ac.bd

Abstract. Load balancing is a crucial element of any distributed system because it serves to spread workloads equally throughout numerous computing resources aiming to optimize speed and efficiency. However, traditional load-balancing techniques often consume a significant amount of energy, which can lead to increased operating costs and negative environmental impacts. We propose a dynamic round-robin approach for cloud load balancing in this paper to alleviate the strain on wireless networks. Our algorithm achieves 8% more efficiency compared to the traditional round-robin model. Additionally, our load-balancing virtual machine architecture exhibits an improvement in resource usage efficiency ranging from 15 to 42% compared to the current model. Overall, our model significantly increases the efficiency of the current load-balancing method.

Keywords: Load balancing · Virtualization · Round robin · Cloud computing

1 Introduction

Data is regarded as a crucial element in the modern day. Data was once managed and stored inside a local computer [1]. As a result of the development of cloud computing, users have the ability to use a collection of adjustable resources that are available for sharing. (Similarly, services, storage, applications, servers, and networks) on a pay-per-use basis [2]. Numerous servers in a data center use several megawatts of electricity each. By distributing data, fog computing seeks to bring data closer to the end consumer. It could almost eliminate bottlenecks brought on by big mass data and increase local storage efficiency. A fog computing setup employs virtual machines (VMs) to form multi-edge devices. Because the number and capacity of edge devices are relatively low, [3–5] fog nodes may change their status by joining, exiting, updating, and so on. The optimal utilization of hardware resources is aided by both the nodes within the system and the efficient service. A load assessment involves identifying a server that is either overloaded or underloaded, as the two distinct phases of this process. A server that is overloaded means it is handling many more requests than it can effectively handle. While

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 299–311, 2024.
https://doi.org/10.1007/978-3-031-50151-7_29

adhering to service level agreements [6–9], the method redistributes the entire workload between overloaded and underloaded servers (SLA). Load balancing is a technique that involves the rerouting of proxy and dispersal of network traffic to various endpoints. Two commonly utilized approaches to load balancing include distributed load balancing and centralized load balancing. The essential elements that are responsible for making [10–15] burden adjustment effective are a backup plan in case some part of the system fails. Energy consumption is a major concern in the wireless networking industry. As more devices become connected to the network, the amount of energy required to sustain them also rises. Many of these devices are often left on and connected to the network for extended periods, increasing energy consumption. There are several factors that can contribute to unnecessary energy consumption, such as slow response rate, starvation, and idle time of devices. To address this issue and optimize energy usage, it is important to consider these factors and identify the most efficient load-balancing methods and techniques. Some of the key questions to consider in this regard include: which existing load balancing method would be the most efficient for optimizing energy usage, and which techniques could be used to make the existing method more efficient. Some of the key objectives in this regard include: proposing a greener distributed load-balancing technique using virtual machines and mathematically validating the proposed load-balancing model.

The structure of the paper is as follows: The introduction to the research paper is included in Sect. 1. Section 2 provides a concise summary of load balancing and different models. Previous research on load balancing is covered in Sect. 3. Section 4 details the method used in the research. Section 5 explains the architecture. In Sect. 6, the implementation, simulation, and performance evaluation of the proposed algorithm and architecture using Docker, Nginx, Golang, Node.js, and Autocannon are discussed, along with the algorithm's features. Finally, Sect. 7 concludes with the paper's conclusion and suggestions for future work.

2 Related Work

This section centers around the prior research and related work on load balancing. Previous research has primarily focused on investigating how to utilize load balancing to enhance energy efficiency.

In [1], Mell and Grance's work offers a comprehensive examination of the characteristics, service models, and deployment models of cloud computing and serves as a helpful reference for understanding the basic concepts of this technology.

In [2], Singh et al. conducted a thorough analysis of the fog computing architecture and its applications in edge computing and big data processing. The authors present a comprehensive review of the current state of fog computing and discuss the various challenges and opportunities in this emerging field.

In [3], Rahmani et al. published a comprehensive overview of the principles, technologies, and applications of fog computing in the context of the Internet of Things.

In [4], Markakis et al. published a comprehensive overview of the emerging advances and applications of cloud and fog computing in 5G mobile networks and discuss the key challenges and opportunities of these technologies in this context. The authors present a detailed analysis of the various enabling technologies for cloud and fog computing in 5G mobile networks, such as software-defined networking, network function virtualization, and edge computing, and consider their potential impact on the performance and efficiency of 5G mobile networks.

In [5], Swathy et al. introduced the SGMLB (Stochastic Gradient Method based Load Balancer) model, which uses a game theoretical approach to dynamically balance the load in cloud environments.

In [6], Mahmood published a comprehensive overview of the concepts, frameworks, and technologies of fog computing. The author discusses the key principles and challenges of fog computing and presents a detailed analysis of the various enabling technologies, such as virtualization, orchestration, and edge analytics used to build fog computing systems.

In [7], Buyya and Srirama discuss the principles and paradigms of fog and edge computing. They describe the key concepts and challenges of these technologies and present a detailed analysis of the enabling technologies used to build fog and edge computing systems. The authors also provide a review of the current state of fog and edge computing research and discuss the various applications of these technologies. However, the accuracy level of their proposed techniques is not high enough.

In [8], Shekhar et al. introduced INDICES, a system that uses principles of Dynamic Data-Driven Application Systems (DDDAS) for performance interference-aware cloud-to-fog application migration. The authors describe the design and implementation of INDICES and demonstrate its effectiveness through simulations and experiments. They also discuss the various challenges and limitations of the system and propose directions for future work.

In [9], Nkenyereye et al. present a study on the architecture, applications, and virtual machine migration of software-defined vehicular cloud networks. The authors describe the design and implementation of these networks and discuss their potential applications in various scenarios. They also present a detailed analysis of the various enabling technologies, such as virtualization, orchestration, and edge analytics, that are used to build these networks.

In [10], Osanaiye et al. reviewed the current state of research on the transition from cloud to fog computing and proposed a conceptual live virtual machine (VM) migration framework. The authors discussed the various challenges and opportunities of fog computing and presented a detailed analysis of the various enabling technologies, such as virtualization, orchestration, and edge analytics used to build fog computing systems. They also reviewed the current state of fog computing research and discussed the various applications of these technologies. The authors also proposed a conceptual live VM migration framework that enables the seamless transition of VMs from the cloud to the fog and discussed its potential benefits and limitations.

In [11], Singh et al. present a study on dynamic task scheduling using a balanced virtual machine (VM) allocation policy for fog computing platforms was proposed. The authors described the design and implementation of this policy and demonstrated its effectiveness through simulations and experiments. They also discuss the various challenges and limitations of the policy and suggest directions for future work. The authors argue that their approach can improve the performance and efficiency of fog computing platforms by dynamically allocating tasks to VMs in a balanced manner.

In [12], Rao and Sree reviewed the current state of research on fog computing, including a conceptual live virtual machine (VM) migration framework, the issues and challenges faced by these systems, and their potential applications. The authors provided a comprehensive overview of the key concepts and technologies of fog computing and discussed the various challenges and opportunities of these systems. They also reviewed the current state of fog computing research and discussed the various applications of these technologies in different domains. Overall, this work is a valuable resource for researchers and practitioners interested in fog computing and its applications.

In [13], Baccarelli et al. conducted a study on the use of fog computing to support energy-saving live migration of virtual machines (VMs) over multipath TCP/IP 5G connections with delay constraints. The authors described the design and implementation of their approach and demonstrated its effectiveness through simulations and experiments. They also discuss the various challenges and limitations of their approach and suggest directions for future work. The authors argue that their approach can improve the energy efficiency of live VM migration over 5G connections by leveraging the capabilities of fog computing.

In [14], Khattak et al. conducted a study on the utilization and load balancing of fog servers for health applications. The authors describe the design and implementation of their approach and demonstrate its effectiveness through simulations and experiments. They also discuss the various challenges and limitations of their approach and suggest directions for future work. The authors argue that their approach can improve the performance and efficiency of fog servers for health applications by dynamically balancing the load across the servers. There is also significant contribution in [16–25].

3 Load Balancing

A load balancer refers to either a hardware appliance or software application that equitably distributes network traffic among a cluster of servers. It dispatches client requests to different servers while monitoring the workload on several active servers. Load balancing also facilitates the addition or removal of servers and guarantees high availability. A generic load-balancing architecture is shown in Fig. 1.

3.1 Types of Load Balancing

- **Static load balancing**: This form of load balancing involves assigning tasks to processors after their creation, with performance evaluation taking place at the outset of execution. It is not considered portable.
- **Dynamic load balancing**: The indicated load balancing method entails routing incoming requests to servers in a manner that prioritizes the server with the least ability to manage workload to receive the load. This can be accomplished dynamically, indicating that the allocation of requests is altered in real-time to avoid any server from becoming overburdened.

3.2 Load Balancing Techniques

- **Centralized load balancing**: Widely adopted load-balancing approach used for IoT technology-based devices by researchers. Fog computing incorporates fog nodes, which enable fog controllers to efficiently manage traffic between networks using centralized load balancing. Because all load-balancing methods are overseen by a central node, they can be easily maintained and swiftly repaired if any issues arise.
- **Distributed load balancing**: In a fog environment, decentralized load balancing means that there is no single node that oversees the computing tasks of all nodes. Instead, each node makes decisions based on its own observations and system knowledge. Such load balancing can be advantageous in a fog computing environment, as it can help users regulate network load.
- **Round Robin**: Fog computing employs numerous homogenous servers that are indistinguishable from one another and offer the same services. Each server on the network has a distinct IP address and is established under the same Internet domain name. This load-balancing method enables straightforward comprehension and implementation of rudimentary fault tolerance capabilities.
- **Source IP Hash**: To handle an incoming request, the source IP address is subjected to a hash function. If the IP hash calculation includes no components, the outcome will be the same. This technique is advantageous because it enables customers to continue their interaction with an active session even after a disconnection or modification.
- **Least response time method**: To guarantee that the end client receives a prompt answer, the backend server with the shortest response time is chosen. When set up to employ this strategy, a load balancer chooses the fewest dynamic connections and the fastest normal response time. The term “time to the first byte” refers to this moment.

4 Method

In a static load balancing approach, round robin is the most widely used algorithm. Round robin is a very simple algorithm to distribute network load among a group of computers in a network. In this method of load balancing, the network requests are routed between the available servers on a cyclic basis. This algorithm is best suited for a group of servers that have the identical computing power and response time.

The way the round-robin algorithm is implemented usually is, the requests are distributed in the same order as they are received. Let’s assume in a network, there are 3 servers which are labeled 1, 2 & 3 shown in Fig. 1. The order in which the requests are distributed is

- The initial request is transmitted to server 1.
- The subsequent request is conveyed to server 2.
- The third request is dispatched to server 3.
- The fourth request is returned to server 1.

Round robin method also has many variations of this algorithm. One of the ways round-robin algorithms are improved is by using weighted round-robin algorithms. In this approach, each server is assigned a node weight which generally defines how much load can this server take compared to the other nodes in this network. For example, let’s assume

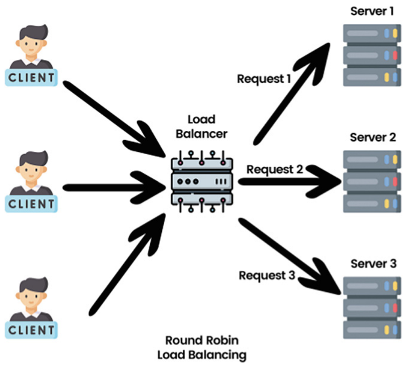


Fig. 1. Round robin load balancing.

- On average, server 1 has the ability to handle 30 requests per second.
- On average, server 2 has the capability to handle 20 requests per second.
- On average, server 2 has the capability to handle 10 requests per second.

The cyclic distribution of requests according to their weight would be done by the weighted round-robin algorithm in this situation. For instance, if there were 6 requests in a row, the algorithm would consider the weight while distributing them in the following order

- Server 1 would be assigned 3 requests.
- Server 2 would be assigned 2 requests.
- Server 3 would be assigned 1 request.

Our proposed load-balancing method takes a dynamic approach to this round-robin algorithm. This proposed algorithm not only considers the order of the requests but also considers several other factors for the distribution such as - the least concurrent connections & time to receive the last byte.

The dynamic round-robin algorithm would take the following steps to determine the order of the request distribution.

1. Store data about each node in a global store, such as - active concurrent connections, time to receive the last byte, and average response time.
2. Sort the server nodes by concurrent connections as the primary key, average response time as the secondary key, and time to last byte as the final key.
3. Cache the sorted result for the next N number of requests and only recompute the order for every N request. (N could vary based on the average network traffic)

Pseudocode

```
active_connections = []
avg_response_time = []
avg_ttlb = []
sorted = []
# Main thread
```

```
  for conn in connections:
    active_connections[server] += 1

     time.begin()
     process_connection(sorted[0], conn)
     avg_response_time[server] = time.end()
     avg_ttlb[server] = time.now()

 # New thread (cached)
 while True:
     sorted = sort(active_connections, avgresponse_time,
avg_ttlb))
     time.sleep(N)
```

Following the steps above would give us a sorted list of the server nodes which are assigned a weight dynamically based on the order of the list. This weight would be changed in the runtime depending on the current server load. Then we can take a weighted round-robin approach to assign the incoming network request to the server nodes.

5 System Architecture

Energy efficiency during load balancing in cloud computing heavily relies on how effectively we are utilizing the available system resources. To make our load balancing technique, we need some significant changes to our system architecture as well. Generally, cloud computing devices are designed to target the highest network load possible. But servers might not always experience the heavy load that the servers were initially designed for. Therefore, server resources are wasted when the network load is on the lower end. It is also difficult to scale up or down physical servers in a resilient manner. If physical servers were to scale down on the runtime, the network might experience brief downtime. To solve this problem, we are going to propose a model utilizing virtualization technology. Instead of scaling server nodes using multiple physical servers, we can scale using containerization technology. This will allow us to scale a server into multiple nodes within the same physical server instance.

Containerization technologies such as Docker or Kubernetes allow us to spin up multiple nodes of the same server that are isolated within their own virtual container. Using this technology, we can scale up our server to as many nodes as needed during the heavy server load that is within the same physical server. This will make it so each physical server is being used up to its full potential before we need to scale the number of physical instances. We can also scale down the number of VMs that we run at a given time if the server loads are on the lower side. Scaling down will lower resource utilization and eventually lower the consumed energy.

Figure 2 demonstrate how we can utilize virtualization and containerization technology to scale up or down our load balancing for energy efficiency. In Fig. 2(a), we can see three physical server instances where physical instance 2 is experiencing congestion due to heavy load. Meanwhile, physical instance 3 is only hosting a single instance of VM wasting resources. In this situation, it is possible to conduct a live migration of a virtual machine and move a VM container from physical server 2 to physical server 3. As a result of the migration shown in Fig. 2(b), all three servers are now running at an equal level, making use of their maximum resources.

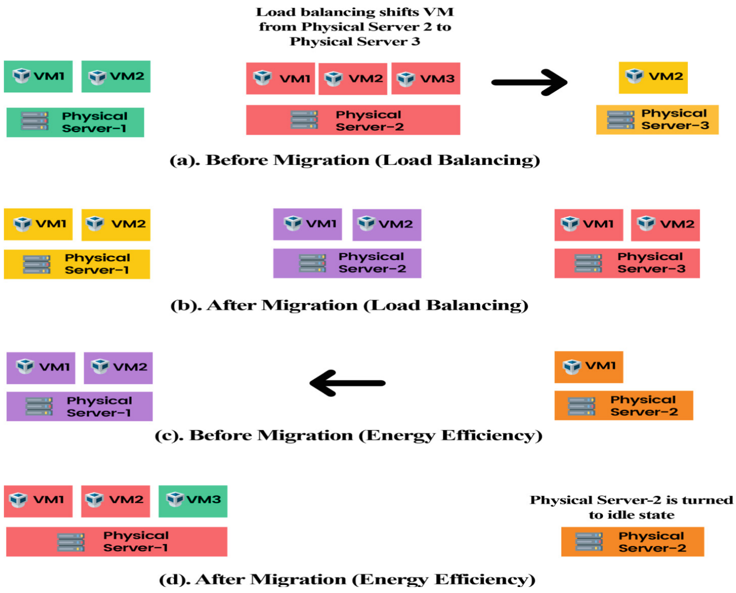


Fig. 2. Live VM migration.

In Fig. 2(c), we can see an opposite case where we have 2 physical server instances and both are operating at a lower capacity than they were originally designed for. This is a case where we can scale down the nodes to only a single physical server instance and perform a live VM migration to scale down our utilized resources. Running a single instance of our physical server will be much more energy efficient since we are able to put physical server 2 into a completely idle state saving energy. In Fig. 2(d), we can see the state after live VM migration of scaling down the number of VM instances.

6 Simulation and Result Analysis

The simulation can be broken down into two parts. The first part will involve comparing the proposed load-balancing algorithm to a round-robin load-balancing algorithm. The second part will involve measuring the resource usage by virtual machines on a larger scale.

For our benchmarking purposes, we are using the following tools

- **Docker**: for containerization & virtualization of server nodes.
- **Nginx**: as our primary load balancer & reverse proxy.
- **Golang**: choice of language for implementing load balancing algorithm.

- **Node.js**: choice of language for serving simple web servers.
- **Autocannon**: HTTP client for stress testing.

Server configuration:

- **CPU**: Apple M2 - 4 core 8 threads (arm64).
- **RAM**: 8GB Unified memory.
- **NIC**: 1024 concurrent connections.

Client configuration:

- **Connections**: 250 to 1000 concurrent connections.
- **Pipeline factor**: 10 pipelines.
- **Worker threads**: 2 threads.
- **Duration**: 15 s.

For benchmarking our dynamic load balancing algorithm, we used 4 instances of our server node and tested both algorithms from the lowest of 250 concurrent connections up to 1000 concurrent connections. For each test case, we measured the average response time in milliseconds. The results of the tests are in Fig. 3.

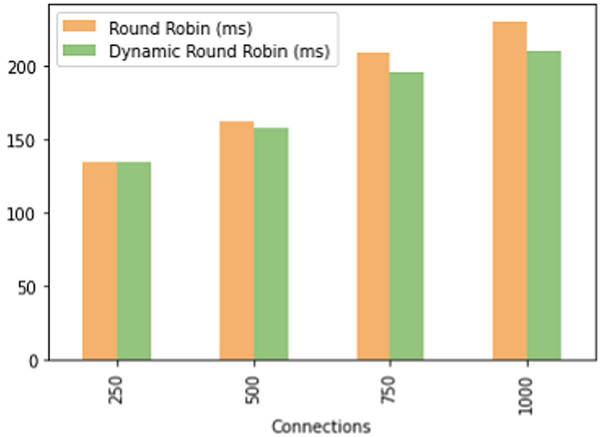


Fig. 3. Round robin versus dynamic round robin (proposed)

From the result, we see response time improvements across different scales. At 250 concurrent connections, we see an identical response time between both algorithms. But the proposed algorithm performs better on a higher scale with up to 8% improvement in response time. The performance improvement using the proposed algorithm in different scales is listed in Table 1.

For the benchmark of the virtualized system architecture, we used dockerized containers scaling from 1 to 4 virtual server nodes. We used our proposed method as a load-balancing algorithm and tested from 250 up to 1000 concurrent connections with

Table 1. Improvement in the dynamic round robin (proposed)

Connections	Round robin (ms)	Dynamic round robin (proposed) (ms)	Improvement (%)
250	135	135	0
500	162	158	2.47
750	209	196	6.22
1000	231	211	8.66

each configuration. For each test case, we recorded the average response time in milliseconds and also the maximum CPU usage during the stress test. The results for the average response time for each configuration are in Fig. 4.

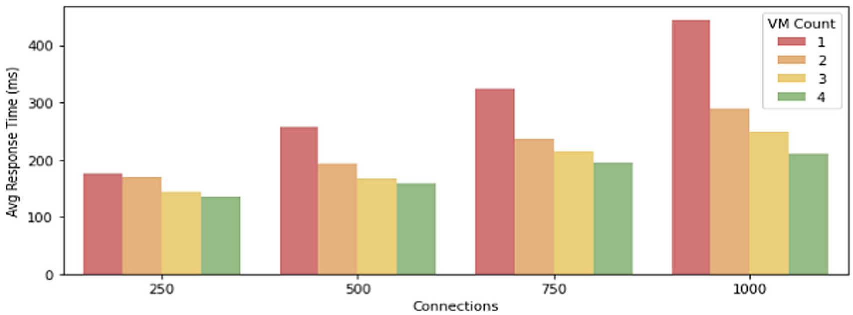


Fig. 4. Average response time (in ms)

The results for maximum CPU usage for each configuration are in Fig. 5.

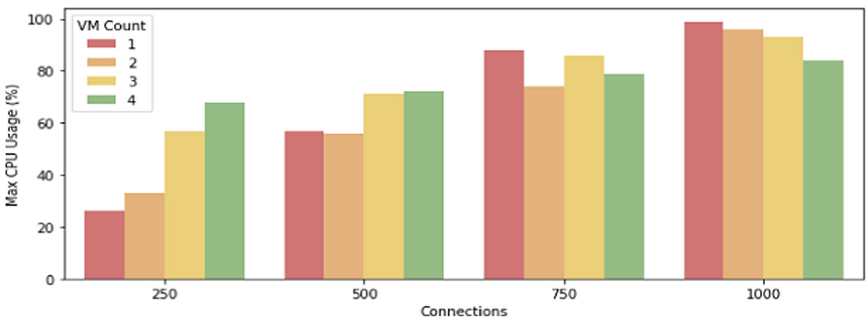


Fig. 5. Max CPU usage

From the result analysis, we see it has significantly lower resource allocation on the lower network traffic since we only run a single virtualized instance of the server node. Whereas on the higher network traffic, we get the complete opposite result. We are able to achieve a 15% better efficiency at 1000 concurrent connections and up to 42% more efficiency in CPU allocation on the lower traffic. This result proves our load balancing method to be effective at reducing energy consumption at load balancing on cloud computing for green wireless networks.

7 Conclusion and Future Works

The use of virtualized nodes in dynamic round-robin load balancing algorithms has been found to be highly effective in cloud computing. Our study introduced a novel approach to this method, which demonstrated an 8% improvement in response time and a 15–42% increase in resource utilization efficiency compared to existing round-robin and physical server methods.

In the future, we can focus on detecting normal traffic and heavy traffic using autonomous systems that can scale up or down our virtualized instances. We can also integrate machine learning into this system to identify patterns in daily network traffic and configure the system architecture based on a cyclic pattern. This prediction method can significantly improve our system's resilience in the long term.

References

1. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. Special Publication 800-145, Gaithersburg (2011)
2. Singh, S.P., Nayyar, A., Kumar, R., Sharma, A.: Fog computing: from architecture to edge computing and big data processing. J. Supercomput. **75**(4), 2070–2105 (2018). https://doi.org/10.1007/s11227-018-2701-2
3. Rahmani, A.M., Liljeberg, P., Preden, J.-S., Jantsch, A.: Fog Computing in the Internet of Things: Intelligence at the Edge. Springer, New York (2017)
4. Markakis, E., Mastorakis, G., Mavromoustakis, C.X., Pallis, E.: Cloud and Fog Computing in 5G Mobile Networks: Emerging Advances and Applications. Institution of Engineering and Technology, London (2017)
5. Swathy, R., Vinayagasundaram, B., Rajesh, G., Nayyar, A., Abouhawwash, M., Abu, E.M.: Game theoretical approach for load balancing using SGMLB model in cloud environment. Plos One **15**(4), e0231708 (2020)
6. Mahmood, Z.: Fog Computing: Concepts. Frameworks and Technologies. Springer, New York (2018)
7. Buyya, R., Srirama, S.N.: Fog and Edge Computing: Principles and Paradigms. Wiley, Hoboken (2019)
8. Shekhar, S., Chhokra, A.D., Bhattacharjee, A. et al.: INDICES: applying DDDAS principles for performance interference-aware cloud-to-fog application migration (2020)
9. Nkenyereye, L., Nkenyereye, L., Tama, B.A., Reddy, A.G., Song, J.S.: Software-defined vehicular cloud networks: architecture, applications and virtual machine migration. Sensors **20**(4), 1092 (2020)
10. Osanaiye, O., Chen, S., Yan, Z., Lu, R., Choo, K.-K.R., Dlodlo, M.: From cloud to fog computing: a review and a conceptual live VM migration framework. IEEE Access **5**, 8284–8300 (2017)
11. Singh, S.P., Nayyar, A., Kaur, H., Singla, A.: Dynamic task scheduling using balanced VM allocation policy for fog computing platforms. Scalable Comput. Pract. Exp. **20**(2), 433–456 (2019)

12. Rao, Y.S., Sree, K.B.: A review on fog computing: conceptual live VM migration framework, issues, applications and its challenges. Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. **3**(1), 1175–1184 (2018)
13. Baccarelli, E., Scarpiniti, M., Momenzadeh, A.: Fog-supported delay-constrained energy-saving live migration of VMs over multipath TCP/IP 5G connections. IEEE Access **6**, 42327–42354 (2018)
14. Khattak, H.A., et al.: Utilization and load balancing in fog servers for health applications. EURASIP J. Wirel. Commun. Netw. **2019**(1), 1–12 (2019). https://doi.org/10.1186/s13638-019-1395-3
15. Singh, S.P., Kumar, R., Sharma, A., Nayyar, A.: Leveraging energy-efficient load balancing algorithms in fog computing. Concurrency Comput.: Pract. Experience **34**(13) (2020). https://doi.org/10.1002/cpe.5913
16. Yeasmin, S., Afrin, N., Saif, K., Reza, A.W., Arefin, M.S.: Towards building a sustainable system of data center cooling and power management utilizing renewable energy. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_67
17. Liza, M.A., Suny, A., Shahjahan, R.M.B., Reza, A.W., Arefin, M.S.: Minimizing e-waste through improved virtualization. In: Vasant, P., Weber, GW., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_97
18. Das, K., Saha, S., Chowdhury, S., Reza, A.W., Paul, S., Arefin, M.S.: A sustainable e-waste management system and recycling trade for Bangladesh in green IT. In: Vasant, P., Weber, GW., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_33
19. Rahman, M.A., Asif, S., Hossain, M.S., Alam, T., Reza, A.W., Arefin, M.S.: A sustainable approach to reduce power consumption and harmful effects of cellular base stations. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_66
20. Ahsan, M., Yousuf, M., Rahman, M., Proma, F.I., Reza, A.W., Arefin, M.S.: Designing a sustainable e-waste management framework for Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_104
21. Mukto, M.M., Al Mahmud, M.M., Ahmed, M.A., Haque, I., Reza, A.W., Arefin, M.S.: A sustainable approach between satellite and traditional broadband transmission technologies based on green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_26
22. Meharaj-Ul-Mahmmud, Laskar, M.S., Arafin, M., Molla, M.S., Reza, A.W., Arefin, M.S.: Improved virtualization to reduce e-waste in green computing. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_35
23. Banik, P., Rahat, M.S.A., Rafe, M.A.H., Reza, A.W., Arefin, M.S.: Developing an Energy cost calculator for solar. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in

Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_75

24. Ahmed, F., Basak, B., Chakraborty, S., Karmokar, T., Reza, A.W., Arefin, M.S.: Sustainable and profitable IT infrastructure of Bangladesh using green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_18
25. Ananna, S.S., Supty, N.S., Shorna, I.J., Reza, A.W., Arefin, M.S.: A policy framework for improving e-waste management in Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_95

Green Cloud Computing: A Sustainable Energy-Efficiency Approach for Business Rapidity and the Environment

Sadia Sultana Bristy[1], Tooba Azam[1], Mukta Marion Islam[1], Raisur Rahman[1], Ahmed Wasif Reza[1](✉), and Mohammad Shamsul Arefin[2,3](✉)

[1] Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
wasif@ewubd.edu

[2] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
sarefin@cuet.ac.bd

[3] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh

Abstract. Cloud computing is significantly assisting in addressing IT issues. To lessen the environmental impact, a thorough investigation of cloud computing and energy-efficient algorithms is required. A green cloud computing strategy aims to build an environment for cloud computing that is more energy-efficient, maximizes resource consumption without overprovisioning, leaves fewer carbon footprints, reduces Greenhouse Gas emissions, and minimizes E-wastages. Some features and techniques that Green Cloud Computing focuses on bringing to the cloud environment include containerization, designing green data center facility infrastructure, server consolidation, virtual machine placement, designing green networks and protocols, green data management, sustainable software development, and safe disposal techniques with the three R's (Reuse, Reduce and Recycle). This paper offers a thorough examination of green cloud computing and its traits. The main objective of this study is to investigate green cloud computing, including its various components and conversion processes, as well as to compare energy-efficient VM placement techniques for green cloud computing.

Keywords: Cloud computing · Green cloud computing · E-waste management · Energy efficiency · Sustainability

1 Introduction

A network of servers that are able to be accessed that have a service to their consumers is referred to as a "cloud." The term "green "refers to living in an eco-sound way that helps to the retention of natural balance and the earth's natural resources [1].

Green cloud computing corresponds to the environmentally sustainable use of computers and their IT-related tools and facilities. It offers detailed fine-grained modeling of the energy consumed by the data center's IT equipment, such as computing servers,

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 312–327, 2024.
https://doi.org/10.1007/978-3-031-50151-7_30

network switches, and communication links. Power management and energy efficiency and lowering carbon footprints are the main goals of green computing. Green computing enhances the way computers are utilized by achieving practicality in terms of cost, environmentally friendly production methods, disposal and recycling techniques, and the usage of energy-efficient resources [2]. Therefore, it is crucial to adhere to green computing principles while building data centers. Additionally, the advantages of sustainable cloud computing are far more significant for preserving the environment.

In this paper, there is a detailed introduction to green cloud computing, its services and different types of cloud models, and essential Characteristics, features, and benefits of Green Cloud Computing have been discussed briefly. There is a comparison of energy-efficient algorithms and the best algorithm is used for green cloud computing architecture. Furthermore, the challenges of green cloud computing are highlighted here.

2 Literature Review

Analyzing data as part of our investigation on "A Sustainable Energy-Efficient Approach for Business Rapidity and the Environment" We carefully go through various journals, conferences, and articles in order to compile the most recent knowledge on the properties of green cloud computing. In the part about the literature, we come up with a review of related prior research articles on green cloud computing.

The paper [3] gives ideas about green computing's characteristics. It also describes the characteristics of green cloud computing like Energy efficiency, Virtualization, Eco Friendly, Consolidation, and Multi-Tenancy. A single physical system can function as numerous systems thanks to virtualization, which effectively allows multi-tenancy.

This research [4] was to examine the dynamics of green cloud computing. This study's starting point was to emphasize the importance of Green Cloud Computing as a strategy for moving forward with sustainable development. This is a relatively new topic in the IT industry, and it must be acknowledged that the current discourse in the field does not emphasize the importance of shifting away from non-renewable energy sources and toward more sustainable ones. The report then discusses key terminology used in the Green Cloud Computing industry, such as PUE and GCA.

In this paper [5], The characteristics of the green computer of the future include efficiency, manufacturing and materials, recycling, a service model, empowerment, and other developments. They offered a few VM migration techniques and architecture.

Another paper [6] provides an in-depth analysis of green cloud computing by elaborating on it. With a quick talk, they covered a variety of topics like energy efficiency, data center power management, and virtualization. A quick explanation of the challenges and global benefits of green cloud computing and an annual evaluation of green IT for the cloud is then presented, together with the perspectives and findings of some of the contributors. The report concludes that the government is taking energy efficiency and power management seriously and setting clear standards as it moves toward a green revolution soon.

In the study [7], they provided an outline of how much energy is used in the cloud computing sector and contrasted it with that of conventional energy sources giving details about the amount of energy used for switching, transmission, data processing, and storage. Additionally, they explored and showed real-world GCC applications, while shedding light on obstacles and potential directions for future GCC technology development research.

In this work [8], a special New Linear Regression (NLR) prediction model, host overload/underload, and VM placement strategy have been advised in order to reduce EC and SLAV. The primary goal of the NLR model is for it to pass through a mean point and a straight line. The proposed NLR model predicts future CPU consumption. The suggested strategies were evaluated using the extended Cloud Simulator.

In a different study [9], an algorithm for optimizing strong agile response task scheduling is suggested based on the duration of job scheduling and the data center's peak energy usage. From the standpoint of task failure rate, the suggested approach may be used to examine the strong agile response optimization model, investigate the probability density function of the task request queue overflow, and submit a timeout request to avoid network congestion.

In the study [10], Green Computing is a new strategy that was created to make data centers more energy-efficient and environmentally beneficial. In order to move toward optimal power usage, as mentioned in advanced green computing research, work is recommended as Power Nap, i.e., an energy-conservation strategy in which servers are moved between active and almost completely power-inactive states idle.

In this study [11], by putting co-related virtual machines (VMs) on the same server, the suggested algorithm, named the virtual machine placement and traffic configuration algorithm (VPTCA), intends to lower the energy usage of data centers. In doing so, data centers all around the world will be able to reduce their transmission burden and increase their energy efficiency.

This study [12] focuses on lowering energy consumption and speeding up work completion. This paper suggests a task-scheduling method that dynamically fuses ant colony and genetic algorithms. Thus reducing the energy consumption of cloud computing's computers and data centers. The work in [27–36] focuses in different areas of green computing.

From the above research, we can say that green cloud computing has a much more impact on our environment and economy. If we can shift to a fully green cloud system, it will reduce our energy consumption because a huge part of the energy is used in Data Centers. The issue of on-site data centers using a lot of electricity is solved by virtualization. It will also reduce CO_2 emissions.

3 Cloud Computing

Utilizing computer system resources as needed for networking, data storage, development, etc. is known as cloud computing [13]. Using the hardware and operating system software found in data centers along with apps that are offered as Internet services makes up cloud infrastructure. All the information that users require access to be in the cloud or virtual environment. Cost savings, adaptability, accessibility, speed, effectiveness, and data security are just a few benefits of cloud computing.

Cloud Computing Deployment Models, Services, and Platforms: With the help of technical advancements, the cloud is a business model that enables consumers to use networks, applications, storage, and computing on a per-use basis without investing in underlying infrastructure. Self-service, elastic, scalable scaling, and pay-per-usage are three separate features of cloud consumption.

3.1 Deployment Models in Cloud

Public Cloud. Cloud resources are made dynamically available through the web utilizing web services or web applications in this off-site location in response to demand and request. Virtualization occurs at several layers of the software stack and enables multi-tenancy, one of the core components of public clouds [14].

Private Cloud. It is a company's internal cloud computing system. Large data centers are installed inside the institution's confines, so only those employees can safely share the resources. The use of private clouds is advised since public clouds can be accessed from outside of an organization, which could pose a security risk or breach. More security control is available with this type of cloud.

Hybrid Cloud. Hybrid clouds offer the chance and access to resources that are too expensive to maintain. Keeping non-critical business application backups on the public cloud, for instance, will increase availability (HA). The hybrid cloud is perfect in terms of scalability, adaptability, and security.

Community Cloud. Businesses or organizations using the community cloud infrastructure have similar or related requirements. They were created with a specific objective in mind, allowing suitable enterprises to share a cloud environment and cooperate on cooperative projects.

3.2 Service Models in Cloud and Their Benefits

Based on responsibilities, user organization, and service providers, cloud computing can deliver a variety of services [15].

IaaS: "Infrastructure as a Service" is the name of it. IaaS buys physical resources like servers, storage, and networks from outside service providers. Users can access the resources via a web API, which is located in another data center. The expanding policy-driven nature of these services enables IaaS users to increase the automation and orchestration of crucial infrastructure operations.

Benefits:

- Resource Management is simple.
- Prevents the need for internal hardware maintenance.
- Through an Internet connection, remote access is possible.
- It is flexible.

PaaS: Platform-as-a-Services allow for the outsourcing of the runtime, databases, and integration layers as well as the physical hardware infrastructure. This approach is perfect for someone who wants to have control over corporate apps and goods.

Benefits:

- Time is saved on project development.
- Production is quick and swift.
- The platform's implementation can be handled automatically.
- Data from businesses can be protected.

SaaS: To utilize this software-as-a-service, which is provided through the Internet, there is no need to install any program. This service is affordable and available to everybody in the world. Most firms prefer cloud solutions since they have a high rate of sales growth.

Benefits:

- New software solutions provide simple and agile testing.
- Don't require any software installation or maintenance.
- They are also free from infrastructure administration and maintenance and update with software updates

3.3 Cloud Computing Platforms

These are the four most well-known cloud computing platforms: Amazon Web Service, Google Cloud, Microsoft Azure, and IBM Cloud. This platform comprises automation, AI/ML, containers, blockchain, IoT, analytics, and automation. Data storage, IoT, containers, etc.

3.4 Need for Green Cloud Datacenter

While managing multiple applications in a data center, it becomes challenging to supply and allocate resources as necessary in response to changing workloads. Data center resources are typically statically allocated to applications based on peak load characteristics to guarantee performance and isolation. Up until recently, data center deployments were solely concerned with providing high performance, and little thought was given to energy consumption. Cloud service providers should take precautions to avoid a sharp decline in their profit margin as a result of rising energy costs [16]. Through the use of efficient processing and computer infrastructure, green cloud computing seeks to use less energy overall. To guarantee the long-term sustainability of the growth of cloud computing, this is essential. To support Green Cloud computing, data center resources must be managed in an energy-efficient manner.

Benefits of green cloud computing:

- Working remotely across the world helps to keep the environment clean.
- Going paperless by using green technology and cloud computing.
- Energy Consumption is reduced by Green Cloud Computing
- E-waste generation is being reduced.

- The replacement of physical products with virtual alternatives.
- Eliminates producing Greenhouse Gas (GHG) Emissions [17]
- Lessen the buildup of hazardous materials.
- Encourage collaboration and remote working

4 Methodology

4.1 Methodology of Data Collection

We have read numerous research papers and journals that were gathered from reputable online resources in order to gather our data. Green modeling, green equipment, and environmental sustainability were some of the topics and keywords that were concentrated on for the systematic literature review. The majority of the papers that were chosen for inclusion in the study were published between 2017 and 2021, and their use in determining the role of sustainability was also considered. Our research focuses on green data centers, green clouds, energy efficiency, e-waste management by green data centers, environmental sustainability, etc.

4.2 Methodology of Data Analysis

The data gathered from the chosen papers was qualitative. In accordance with their functions and roles, the data were categorized. The information was further examined in order to determine the process for converting cloud computing into green cloud computing, its advantages for business rapidity, as well as its opportunities and difficulties. The targets and goals were divided into categories to determine cloud computing's energy efficiency and turn it into a green cloud center through green modeling for business rapidity and environmental sustainability.

4.3 Techniques to Turn Cloud into Green Cloud

Although there are some significant concerns in society that the use of the cloud increases the amount of energy used in huge data centers and increases carbon emissions, there are some environmentally friendly aspects of cloud computing [18]. By converting the data centers into energy-efficient infrastructure by using the technologies, cloud service providers can achieve improved efficiency and high resource utilization, which results in lower carbon emission rates.

Multi-tenancy. Cloud computing can be energy efficient because multiple firms can share the same hardware and software. Multi-tenancy (Fig. 1) means that the same server can be shared during periods of high business demand, balancing the fluctuating demand patterns. This reduces overall energy consumption and CO_2 emissions, resulting in green cloud computing architecture.

Dynamic Provisioning. CSPs should dynamically assign resources in accordance with application requirements. Distributing virtual machines that are taken from a resource pool enables dynamic provisioning. Using virtual volumes rather than actual disk space,

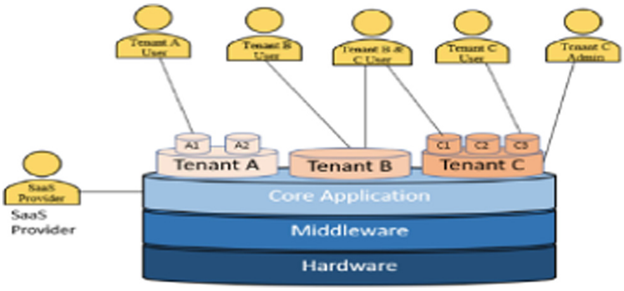


Fig. 1. Multi-tenant application

Dynamic Provisioning enables you to reserve virtual storage capacity depending on expected future capacity needs [19].

Data Center Power Efficiency. The overall power consumption in the data center has a greater environmental impact in cloud computing. A contemporary data center with cutting-edge cooling systems and technologies must be built in order to maintain Power Usage Efficiency (PUE) levels between 1.1 and 1.2. The host design should support bare metal, containers, or virtualization for modern applications with good power efficiency [20].

Host Virtualization. The virtualization method (Fig. 2), which enables more efficient use of actual computer hardware, is the foundation upon which cloud computing is constructed. When programs are executed on a single host using virtualization technologies, they can be utilized at high levels with up to 65% greater performance and with increased energy efficiency. As a result of server virtualization, there are much fewer physical servers in operation.

Fig. 2. Multi-tenancy

Hypervisor Role in Virtualization

Virtualization is made possible by a piece of software called the hypervisor. It divides the operating system (OS) and guest computer from the installed hardware. The forthcoming virtual machines are separated from the CPU, memory, and other hardware resources by the virtualization layer that a hypervisor produces [21].

Application Container in POD

In the CISCO containerized data center (Fig. 3), Power Usage Efficiency (PUE), also known as Performance Optimized Data Center (POD), gradually grew from 1.05 to 1.25. Container startup time is considerably lower than VM startup time since the program in a container is mainly isolated from the OS and the container is application-centric. This results in the containers being lightweight. The containerized data center, which has a compact footprint, can supply the self-contained data center in almost any place**.**

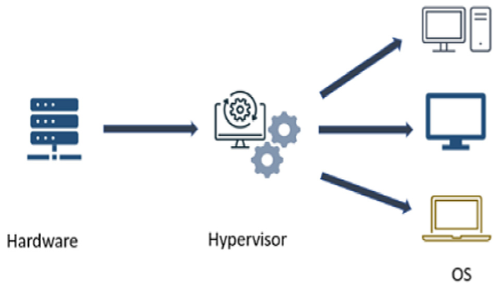


Fig. 3. Hypervisor

5 Green Cloud Architecture

Green brokers can access the public directory. Green offers include green services like cost, timing, and appropriate times to use them for a smaller carbon footprint. This Green Cloud Architecture's (Fig. 4) carbon emission directory includes Carbon Usage Effectiveness (CUE), Power Usage Efficiency (PUE), and other energy indicators. To better serve the user, the green cloud framework monitors overall energy use. From the end user's perspective, choosing green cloud services with Quality of Service (QOS) is challenging; as a result, each cloud layer must be "Green Aware" at the cloud service provider's end [22].

5.1 Green Data Center

Green data centers should be constructed and operated with IT, electrical, and air conditioning systems to promote optimum energy efficiency and minimal environmental effect. For operating and processing requirements, the data center should employ cutting-edge cooling and IT technologies that are energy efficient. The approaches of blade servers, green data storage, cutting-edge cooling concepts, server clustering, and green networks can reduce the energy usage of data centers.

Green Data Storage. Green storage is an iterative process of data and storage system design and execution decisions that reduces the data center storage footprint for data

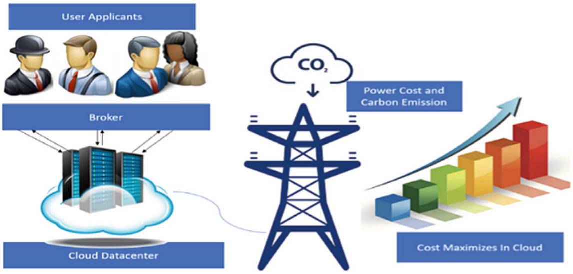


Fig. 4. Architecture of green cloud

center IT operators and storage administrators. The need for storage has substantially expanded as a result of big data, and data center storage systems are consuming more cooling and power [23]. Data deduplication is a technique that eliminates redundant data to improve storage effectiveness, allowing for the saving of only unique data. To eliminate redundant files and data, the tiering technique is applied. Tiering also aids in supplying data instantly for business analytics. Tier-1 is utilized for data that is available immediately, Tier-2 is used for timely but non-critical data, and Tier-3 is helpful for archiving data.

Blade Server. Blade systems are fundamentally different from rack systems. While blade systems are kept inside a container of blade system housed vertically in a typical cabinet, rack systems are self-contained, independent devices that need network cabling and power for each unit. Uninterrupted Power Supply (UPS) and onboard cooling are features of the blade cabinet system. Blade servers are incredibly thin and can be changed out without causing any disruptions. The most energy-efficient server, according to PSSC Labs' new eco-blade model, can reduce physical space by up to 46% while delivering power savings of up to 46% per server. Blade servers offer simple web-based system management.

Green Network and Communication. The practice of choosing energy-efficient networking tools and products while minimizing resource consumption is known as green networking. The on-demand nature of communication networks due to information sharing, online purchasing, and e-learning raises energy usage. Traditional network protocols were not created with energy conservation in mind. It exhibits an adverse effect on the environment. By utilizing eco-protocols, networking technology, and goods, green networking reduces its negative environmental effects.

- Begin using brand-new network products that are energy efficient.
- Begin using power management and smart systems to create IT networks energy-efficient
- Begin upgrading the gadget with an energy-saving protocol scheme.

Cooling Systems in Green Data Center. In the data center (Fig. 5), heat is generated by the majority of the servers and other equipment. The purpose of cooling in the DC is to prevent overheating of information technology (IT) equipment from impairing or destroying performance and resulting in service interruptions. The server rack in the data center faces cold aisles in the front and hot aisles in the back. A room, rack, or row-based system is an air-based system. Systems built on racks are more effective because the power used to supply the airflow is restricted to the rack. When compared to room-based air-conditioning systems, row-based is more effective. Mechanical refrigeration can be replaced in Green Data Centers with set-side, water-side, or chilling to reduce energy usage. The Heat Exchanger (HEX) in wet-side cooling simply returns water, while heat escapes through the dry cooler to the environment outside. In order to reduce cooling costs and maintain environmental sustainability, modern data centers are now built under strict regulations near lakes or rivers for waterside cooling.

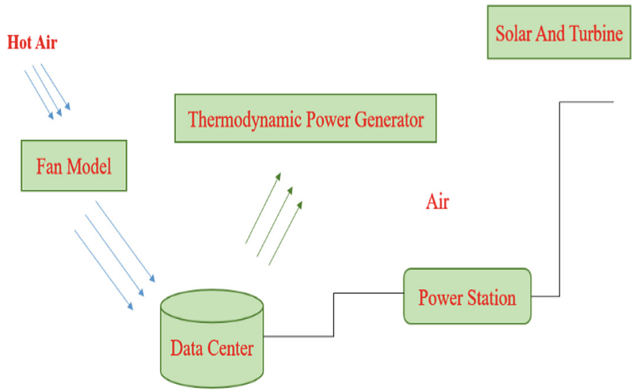


Fig. 5. Green data center's cooling system

Green Software Sustainability. The definition, development, and administration of sustainable software applications are the primary goals of the developing field of "Green Software Engineering.". To control and track power management, data centers use software that is ecologically friendly. To reduce e-waste and to improve the reuse of modules without having to construct them from start again, which again saves time, money, and energy, it should be developed with the reuse of code, agile approach, and portability to run on diverse environments.

5.2 Study of Green Cloud Computing's Energy Conservation

Analyzing energy efficiency in green data centers is crucial. A crucial technology for the cloud computing environment is virtualization. By reducing the amount of energy used in data centers, VM consolidation assigns virtual machines to as few servers as possible. The practicality of the open-source consolidation framework known as OpenStack Neat is what sets it apart from other frameworks. Its components have proven useful for choosing the right host for the right VM and determining when to migrate it. Modified Best-Fit Decreasing is the name of the VM placement algorithmic tool used by OpenStack Neat. These algorithms rely on an analysis that uses less energy and raises service level agreement (SLA) violations since it requires more VM migrations. Then, in order to solve the shortcomings of the current VM placement with regard to elements like energy efficiency, VM migrations, and SLA violations, researchers develop some new VM placement methods. The performance of various algorithms, including the Modified Best-Fit Decreasing (MBFD), Efficient Power First-Fit Decreasing (PEFFD), Efficient Power Best-Fit Decreasing (PEBFD), and Medium-Fit Power Efficient Decreasing (MFPED) algorithms, has been assessed. To assess algorithm performance, we ran tests using a simulator called CloudSim.

5.2.1 Summary of Various Algorithms

There are several VM placement techniques that have been studied by a researcher and have been shown to be efficient in decreasing power usage and SLA violation [24]. The algorithms used here are Power Efficient First-Fit Decreasing (PEFFD) Algorithm, Power Efficient Best-Fit Decreasing (PEBFD) Algorithm, Medium-Fit Power Efficient Decreasing (MFPED) Algorithm, and Modified Best-Fit Decreasing (MBFD) Algorithm.

6 Experimental Setup and Simulation Report

The aforementioned algorithms are simulated (as in Table 1) using the CloudSim simulator, an open-source tool that helps apply the concept of resource allocation in a cloud context. In order to determine the optimum algorithm for VM placement, we have examined algorithm energy usage and overall SLA violation [1].

Table 1. Simulation data

Algorithm	Number of hosts	Number of VMs	Total simulation time(s)	Energy consumption (kwh)	SLA (%)	Overall SLA violation (%)	Average SLA violation (%)
PEFFD	800	1052	86400.00	205.65	0.00215	0.48	14.11
PEBFD	800	1052	86400.00	205.45	0.00213	0.47	14.18
MFPED	800	1052	86400.00	204.84	0.00218	0.43	13.77
MBFD	800	1052	86400.00	246.67	0.00319	0.50	13.24

From the above overall comparison (Figs. 6 and 7), we can see MFPED gives the best energy efficiency and less SLA violation.

7 E-Waste Management in Green Cloud Computing

E-waste impacts both humans and the environment in negative ways. When in use, electronics and electrical equipment appear to be effective, resourceful, and ecologically beneficial, but after they become outdated or e-waste, they come with hidden threats.

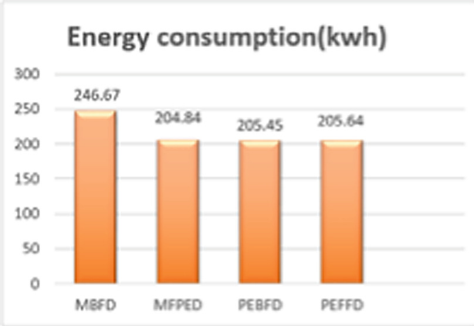


Fig. 6. The compassion of all algorithms with respect to their energy consumption

Fig. 7. The compassion of all algorithms with respect to SLA Violation(in percentage)

Reducing the use of materials that have a detrimental influence on the environment through design and usage is the aim of green computing. The way e-waste can be managed is by following 3R's process [25, 26]. The way e-waste can be managed is by following 3R's process.

Reduce. E-waste can be reduced by implementing virtualization in every possible way. Manufacturing businesses can participate in this initiative by utilizing the public cloud data center services provided by Amazon, Microsoft, HP, and Google, all of which have virtualization technology in their own data centers.

Re-use. Pre-owned electronic devices can be put to good use by other people. Discarded products can be used again by testing their fitness and working capacity.

Recycle. E-waste includes recyclable plastics, bulk materials like iron and aluminum, and precious metals like copper, gold, silver, platinum, and palladium. Recycling products use safe disposal methods, such as sorting the components and classifying them for recycling.

8 Expected Results

Analysis of green cloud computing, here the concept of cloud computing is explored, as the need of designing green clouds. Research has been conducted to address the challenge of energy consumption using the VM consolidation technique. With a comparison of the baseline algorithm called **MFPED**, the algorithms mentioned above are more power efficient and we implement this algorithm to design a green data center. Moreover, if we are successful in implementing these strategies, we will be able to lower energy consumption, e-waste output, and the number of physical products. The environment will be protected by becoming paperless and using green technologies and cloud computing. The environment will be protected by becoming paperless and using green technologies and cloud computing. We can perform global remote work to preserve a clean environment.

9 Challenges

Some of the difficulties green cloud computing faces are covered in this section.

Virtualization. Due to optimization, cloud virtualization has some substantial drawbacks. A crucial research task is developing unique strategies using cutting-edge technology to optimize the whole life cycle of virtualization. The automated development of perfect virtual machines with sufficient resources without impacting cloud performance is another significant problem.

Energy efficiency. Software for data processing and storage in the cloud uses a lot of energy. The cloud needs a strong electrical control system in order to be energy efficient. An integrative and intelligent approach is needed to address energy optimization issues throughout the entire cloud infrastructure layer. It is also vital to have sophisticated systems for distributing energy and making power supply decisions.

Tenancy. Right now, there are protection and privacy problems with multi-tenancy. Designing a protected multi-locator architecture and ensuring that multi-locator modules may be accessed safely present considerable problems.

Consolidation. Potential analysis problems for the virtual machine consolidation concept's information support included managing server downtime, using core resources, and estimating the importance of several aspects of a threshold.

Environmental Friendliness. Designing software with an environmental focus, such as a carbon calculator, to measure the influence of the cloud on the environment is the main goal. There is a significant need to build a strict cloud qualification system that is focused on several aspects of green cloud computing.

10 Conclusion

Green cloud computing is a promising solution for reducing the environmental impact of the IT industry. Accepting green cloud computing is a vital and necessary step toward a future that is more environmentally friendly and sustainable. Businesses and people can use green cloud-based technology to cut down their energy use and carbon emissions while still having access to the computational capacity they require. Cloud service providers can drastically lower their carbon footprint and contribute to a more sustainable future by implementing sustainable practices like the use of renewable energy, green data centers, and others. By doing this, we can make sure that everyone can benefit from cloud computing without harming the environment.

References

1. Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A.F., Buyya, R.: CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Wiley Online Library (2011)
2. What is green computing?—definition from Techopedia
3. Kirvan, P.: How much energy do data centers consume? Techtarget (2022)
4. Sönnichsen, N.: Global data centers energy demand by type 2015–2021. Statista
5. Singha, A., Sarkar, S.J., Nayak, S., Patgiri, R.: Green cloud computing-to build a sustainable tomorrow. IEEE (2022)
6. Gopala, Sriram, G.K.: Green cloud computing: an approach towards sustainability. Researchgate (2022)
7. Sharma, R., Bala, A., Singh, A.: Virtual machine migration for green cloud computing. Researchgate (2022)
8. Shelke, P., Saini, J.K., Agrawal, N.: Review on green cloud computing: a step towards saving global environment. Researchgate (2021)
9. Abd El-Mawla, N., Ibrahim, H.: Green cloud computing (GCC), applications, challenges and future research directions. Nile J. Commun. Comput. Sci. (2022)
10. Biswas, N.K., Banerjee, S., Biswas, U., Ghosh, U.: An approach towards development of new linear regression prediction model for reduced energy consumption and SLA violation in the domain of green cloud computing. Sustain. Energy Technol. Assess. (2021)
11. Shu, W., Cai, K., Xiong, N.N.: Research on strong agile response task scheduling optimization enhancement with optimal resource usage in green cloud computing. ACM (2021)
12. Meisner, D., Gold, B.T., Wenisch, T.F.: PowerNap: eliminating server idle power. ACM (2009)
13. Yang, T., Lee, Y.C., Wang, C., Zhou, B.B., Chen, J., Zomaya, A.Y.: Workload characteristic oriented scheduler for mapreduce. In: Proceedings of the 18th IEEE international conference on parallel and distributed systems (ICPADS), pp. 156–163 (2012)
14. Zong, Z.: An improvement of task scheduling algorithms for green cloud computing. IEEE (2020)
15. Frankenfield, J.: What is cloud computing? Pros and cons of different types of services. Investopedia
16. Bharani Dharan, G., Jayalakshmi, S.: Harnessing green cloud computing—an energy efficient methodology for business agility and environmental sustainability. Int. J. Emerg. Trends Eng. Res. (2020). ISSN: 2347-3983
17. Goyal, V., Garg, S.: Review on green cloud computing for sustainable environment. Proc. Int. J. Artif. Intell. Internet Things Cloud Comput. **1**, 51–57 (2022)
18. Cloud computing service providers. educba
19. Sheldon, R.: Green Cloud. Techtarget
20. Bonuccelli, G.: Learn the environmental and cost benefits of green cloud computing. Parallels blogs
21. What is a Hypervisor? |VMware Glossary
22. About Dynamic Provisioning—Hitachi Vantara Knowledge

23. Murugesan, S., Gangadharan, G.R.: Green data storage. IEEE (2012)
24. Moges, F.F., Abebe, S.L.: Energy-aware VM placement algorithms for the OpenStack Neat consolidation framework. Moges Abebe J. Cloud Comput.: Adv. Syst. Appl. **8**, 2 (2019)
25. Gayathri, B., Anbuselvi, R.: Effect of green cloud computing and environmental sustainability approaches. Proc. Int. J. Emerg. Technol. Comput. Sci. Electron. (IJETCSE) **12**(4) (2015). ISSN: 0976-1353
26. Kaur, M.: Green cloud computing—future of green IT. Proc. J. Emerg. Technol. Innov. Res. (JETIR)
27. Yeasmin, S., Afrin, N., Saif, K., Reza, A.W., Arefin, M.S.: Towards building a sustainable system of data center cooling and power management utilizing renewable energy. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_67
28. Liza, M.A., Suny, A., Shahjahan, R.M.B., Reza, A.W., Arefin, M.S.: Minimizing e-waste through improved virtualization. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_97
29. Das, K., Saha, S., Chowdhury, S., Reza, A.W., Paul, S., Arefin, M.S.: A sustainable e-waste management system and recycling trade for bangladesh in green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_33
30. Rahman, M.A., Asif, S., Hossain, M.S., Alam, T., Reza, A.W., Arefin, M.S.: A sustainable approach to reduce power consumption and harmful effects of cellular base stations. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_66
31. Ahsan, M., Yousuf, M., Rahman, M., Proma, F.I., Reza, A.W., Arefin, M.S.: Designing a sustainable e-waste management framework for Bangladesh. In: Vasant, P., Weber, GW., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_104
32. Mukto, M.M., Al Mahmud, M.M., Ahmed, M.A., Haque, I., Reza, A.W., Arefin, M.S.: A sustainable approach between satellite and traditional broadband transmission technologies based on green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_26
33. Meharaj-Ul-Mahmmud, Laskar, M.S., Arafin, M., Molla, M.S., Reza, A.W., Arefin, M.S.: Improved virtualization to reduce e-waste in green computing. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_35
34. Banik, P., Rahat, M.S.A., Rafe, M.A.H., Reza, A.W., Arefin, M.S.: Developing an energy cost calculator for solar. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_75
35. Ahmed, F., Basak, B., Chakraborty, S., Karmokar, T., Reza, A.W., Arefin, M.S.: Sustainable and profitable IT infrastructure of Bangladesh using green IT. In: Vasant, P., Weber,

G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_18

36. Ananna, S.S., Supty, N.S., Shorna, I.J., Reza, A.W., Arefin, M.S.: A policy framework for improving e-waste management in Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_95

Prediction of Power Consumption and Indoor Ambient Conditions by Artificial Neural Network and Long Short-Term Memory

Tosapon Intaraumnauy, Nattavit Pimavilai, and Somporn Sirisumrannukul(✉)

Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's Univerity of Technology North Bangkok, 1518 Pracharat 1 Rd. Wongsawang, Bangsue, Bangkok, Thailand
{s6401011810056,s6301011910027}@kmutnb.ac.th, somporn.s@eng.kmutnb.ac.th

Abstract. A data analysis model is proposed to predict the indoor ambient conditions and the power consumption of an air conditioner from the information obtained from air conditioner adjustment and ambient conditions. The ambient and usage data were collected from sensors and internet of things (IoT) devices installed in an occupied home. An artificial neural network (ANN) has been developed to predict the power consumption of the air conditioner after adjusting the air conditioner's target temperature, fan mode setting, and ambient conditions. In addition, a long short-term memory (LSTM) has been developed to predict the indoor temperature and the humidity of the air conditioner when it is controlled. With the information on the upcoming indoor ambient conditions that are expected to occur, the data analysis model can also indicate the maximum energy savings while satisfying the constraint of the resident's comfort. The experimental results demonstrate that the proposed data analysis model can reliably predict the ambient conditions in the air-conditioned room and the power consumption of the air conditioner from the air-conditioning adjustment and ambient, giving the best-operating settings in any different environment scenarios and therefore shows potential for home energy savings for smart home applications.

Keywords: Power consumption prediction · Ambient conditions prediction · Artificial neural network · Long short-term memory · Data analytics

1 Introduction

Nowadays, most people can access technology more than ever before with a variety of technological devices widely available and their market prices are many times lower than before and become affordable. As a result, most people have used various technological devices to facilitate their daily lives even more, including the use of those to facilitate home use. Energy management systems in residential homes have become more and more important to improve energy efficiency and are one of the most frequently mentioned areas of research.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 328–338, 2024.
https://doi.org/10.1007/978-3-031-50151-7_31

Households in tropical countries have a higher demand for air conditioners, and household electricity bills are mostly occupied by the electricity consumed by air conditioners. Many research efforts have paid attention to exploring the ways technology devices in the market such as sensor devices and remote-control devices can help improve energy efficiency and reduce electricity bills. Home energy management systems can offer a workable solution for controlling energy consumption to achieve maximum efficiency without affecting the comfort of the user.

A survey of related literature [1, 2] found that most predictions of indoor ambient conditions use the ANN and focus only on the prediction of upcoming room temperatures. And there is no estimate of the power consumption of the air conditioner from its use. However, predicting the upcoming room humidity is very important, as humidity can be a translator for assessing occupant comfort [3]. Also, using the LSTM to predict indoor ambient conditions is more suitable because it is a sequence problem. This research paper proposed a data analytics model for a cooling management system. The model is equipped with two algorithms: an artificial neural network (ANN) and a long short-term memory (LSTM). Given ambient conditions, the ANN predicts the power consumption based on adjusting the air conditioner's target temperature and fan mode setting. The LSTM algorithm predicts the indoor temperature and the humidity of the air conditioner when it is controlled. Field experiments were tested with a real air conditioner, smart meters, temperature and humidity sensors, and infrared bluster.

2 Methodology

The data analysis model is developed with the aim of predicting the power consumption of an air conditioner and the indoor ambient conditions that would occur in the next period after detecting air conditioner adjustment. The ambient data that influence the prediction include indoor temperature, indoor humidity, and outdoor temperature. The air conditioner adjustment affecting the prediction includes target temperature adjustment and fan mode adjustment. The data analysis model for prediction uses an artificial neural network (ANN) combined with a long short-term memory (LSTM). However, the analytical model needs to train both networks first before them to predict outcomes. The architecture of the analytical model is shown in Fig. 1.

2.1 Data Preparation

Data preparation is the first step that brings the collected dataset for sorting, editing, and calculating additional data. Through this process, the provided dataset becomes complete and ready to train the ANN and the LSTM to determine optimal weight values and optimal bias values before both networks are used to predict the power consumption of the air conditioner and the upcoming indoor ambient conditions.

The collected dataset includes six parameters: power consumption ($Power_t$), outdoor temperature ($OutTemp_t$), indoor temperature ($InTemp_t$), indoor humidity ($InHum_t$), target temperature ($TargetTemp_t$), and fan mode of the air conditioner ($FanMode_t$). The power consumption data are collected by a smart meter whereas the data on outdoor temperature, indoor temperature, and indoor humidity are collected by temperature and

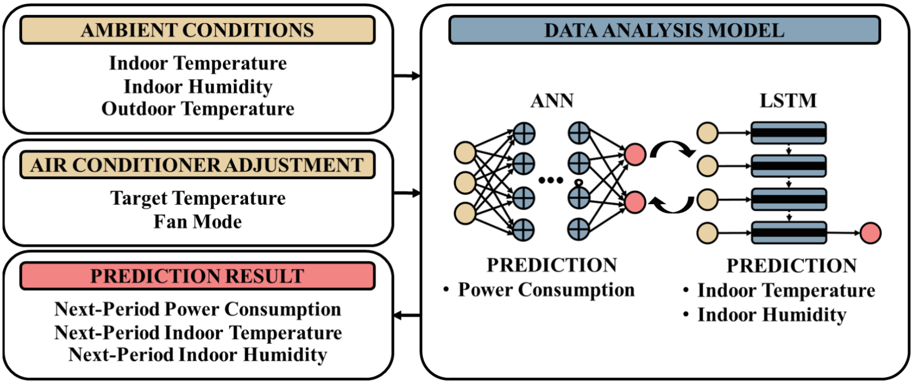


Fig. 1. Architecture of the proposed data analytics model

humidity sensors installed inside and outside the air-conditioned room. The target temperature and fan mode are stored with an IR blaster that acts as a remote control for the air conditioner. The data are stored as datasets at 1-min intervals. An example of the collected dataset is shown in Table 1.

Table 1. Example of collected dataset

Index	$Power_t$ (W)	$OutTemp_t$ (°C)	$InTemp_t$ (°C)	$InHum_t$ (rh)	$TargetTemp_t$ (°C)	$FanMode_t$ (Mode)
t	894	30.25	25.8	49.42	25	2
$t+1$	841	30.25	25.8	49.42	25	2
$t+2$	841	30.25	25.8	49.42	25	2
$t+3$	769	30.25	25.3	49.42	25	2
$t+\mathrm{n}$	769	30.25	25.3	49.42	25	2

In the data preparation process, input is the collected dataset, and output is the input dataset for the ANN. The output dataset for the ANN, the input dataset for the long short-term memory, and the output dataset for the long short-term memory. The data preparation process is given as follows.

Step 1: The collected dataset is linearized for outdoor temperature data, indoor temperature data, and indoor humidity data.

Step 2: The additional data are created. The difference in temperature ($DiffTemp_t$), next-period indoor temperature ($InTemp_{t+1}$), and next-period indoor humidity ($InHum_{t+1}$). The variable $DiffTemp_t$ is calculated from the difference between the indoor temperature and the target temperature. The variable $InTemp_{t+1}$ is set equal to the next value of indoor temperature in the dataset. The $InHum_{t+1}$ is set equal to the next value of indoor humidity in the dataset.

Step 3: Data screening is performed by eliminating some of the input time-series data with a power consumption of fewer than 100 watts because such low-power consumption data are very likely not caused by the operation of the air conditioner.

Step 4: The dataset is divided into four groups: input datasets for the ANN consist of outdoor temperature, indoor temperature, indoor humidity, difference in temperature, and fan mode. The output datasets for the ANN consist only of power consumption. The input datasets for the LSTM include indoor temperature, indoor humidity, and power consumption. The output datasets for the LTSM are next-period indoor temperature and next-period indoor humidity.

Step 5: The input data are normalized for the ANN and the LSTM with the standardized features method by removing the mean and scaling to unit variance [4]. Examples of input datasets for the ANN, output datasets for the ANN, input datasets for the LSTM, and output datasets for the LSTM are shown in Tables 2, 3, 4 and 5 respectively.

Table 2. Examples of input normalized datasets for ANN

Index	$\boldsymbol{OutTemp_t}$ **(Normalized)**	$\boldsymbol{InTemp_t}$ **(Normalized)**	$\boldsymbol{InHum_t}$ **(Normalized)**	$\boldsymbol{DiffTemp_t}$ **(Normalized)**	$\boldsymbol{FanMode_t}$ **(Normalized)**
t	1.5465	1.1426	-0.8199	0.5485	1.1062
$t+1$	1.5441	1.0801	-0.8428	0.4805	1.1062
$t+2$	1.5417	1.0175	-0.8658	0.4125	1.1062
$t+3$	1.5393	0.9550	-0.8888	0.3444	1.1062
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t+\mathrm{n}$	1.5369	0.9237	-0.9117	0.3104	1.1062

Table 3. Example of output datasets for ANN

Index	$\boldsymbol{Power_t}$ **(W)**
t	894
$t+1$	841
$t+2$	841
$t+3$	769
$\vdots$	$\vdots$
$t+\mathrm{n}$	769

Step 6: The input datasets prepared for the LSTM must be compiled into a 3-dimensional array [5, 6]. The first dimension, called Batch, is the number of training examples. The second dimension, called Timesteps, is the number of historical feature

Table 4. Example of input datasets for LSTM

Index	***$Power_t$*** **(Normalized)**	***$InTemp_t$*** **(Normalized)**	***$InHum_t$*** **(Normalized)**
t	2.1644	1.1426	-0.8199
$t+1$	1.9203	1.0801	-0.8428
$t+2$	1.9203	1.0175	-0.8658
$t+3$	1.5887	0.955	-0.8888
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t+\mathrm{n}$	1.5887	0.9237	-0.9117

Table 5. Example of output datasets for LSTM

Index	***$InTemp_{t+1}$*** **(°C)**	***$InHum_{t+1}$*** **(rh)**
t	25.50	49.152
$t+1$	25.40	49.062
$t+2$	25.30	48.973
$t+3$	25.25	48.884
$\vdots$	$\vdots$	$\vdots$
$t+\mathrm{n}$	25.20	48.795

data used for prediction (e.g., 30 min). The third dimension, called Features, is the number of features in each input. In this case, there are three features per timestep: power consumption, indoor temperature, and indoor humidity.

Note that the input and output data set prepared for the ANN and the LSTM is divided into 2 groups: training data set occupying 80% of the data set, and test data set covering the rest of 20% [7]. The input and output of the training data set and the test data set have the same parameters.

2.2 Prediction of Power Consumption by Artificial Neural Network

An artificial neural network is a mathematical model capable of learning [8, 9]. The operation of an artificial neural network is divided into two parts: training and application for prediction. In the training section, the analytical model trains the ANN using the provided data to determine optimal weight values and optimal bias values. In the application section, the ANN is used to predict power consumption from air conditioner adjustment and ambient.

For predicting power consumption ($Power_t$), there are 5 inputs required in the model: indoor temperature ($InTemp_t$), indoor humidity ($InHum_t$), outdoor temperature ($OutTemp_t$), fan mode ($FanMode_t$), and the difference in temperature ($DiffTemp_t$), as shown in Fig. 2.

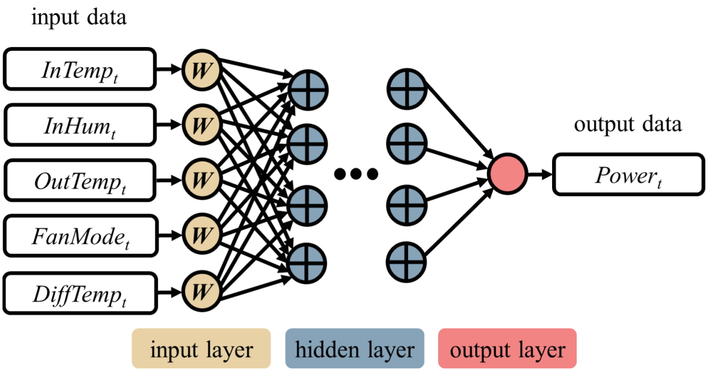


Fig. 2. Artificial neural network for predicting power consumption

2.3 Prediction of Indoor Ambient by Long Short-Term Memory

Long short-term memory (LSTM) is a type of artificial neural network designed for sequence processing [10, 11]. Like the ANN mechanism, the operation of LSTM is divided into two parts: training and application. In the training section, the analytical model trains the LSTM using the provided data to determine optimal weight values and optimal bias values. In the training section, the LSTM is used to predict the indoor ambient conditions that are likely to occur in the next period from the power consumption of the air conditioner and indoor ambient conditions.

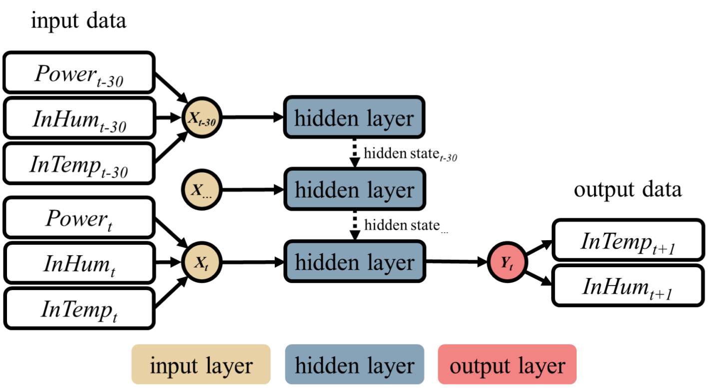


Fig. 3. Long short-term memory for predicting next-period indoor ambient conditions

For predicting the next-period indoor temperature ($InTemp_{t+1}$) and the next-period indoor humidity ($InHum_{t+1}$), there are 3 inputs required: power consumption ($Power_t$), indoor temperature ($InTemp_t$), and indoor humidity ($InHum_t$), as shown in Fig. 3.

2.4 Time-Series Data Analytics Model

The data analytics model combines the two features of the developed ANN and LSTM to predict power consumption, indoor temperature, and outdoor temperature in time series formats. The data analytics algorithm by the ANN and the LSTM is shown in Fig. 4 and Table 6.

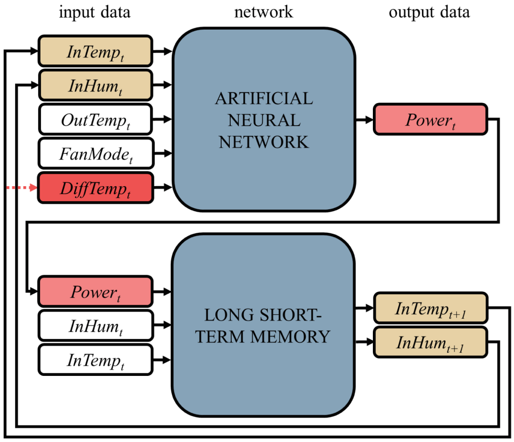


Fig. 4. Prediction model of time-series power consumption and indoor ambient

Table 6. Pseudocode for predicting time-series power consumption and indoor ambient by ANN and LSTM

Input	$InTemp_t$, $InHum_t$, $OutTemp_t$, $TargetTemp_t$, $FanMode_t$,
Output	$Power_t$, $InTemp_{t+1}$, $InHum_{t+1}$
1	**FOR** $t = 0$ to the predetermined period of time
1-1	**IF** $t > 0$
1-1-1	$InTempt = InTempt+1$
1-1-2	$InHumt = InHumt+1$
1-2	$DiffTemp_t = InTemp_t - TargetTemp_t$
1-3	$Power_t$ = prediction results of the ANN from: $InTemp_t$, $InHum_t$, $OutTemp_t$, $FanMode_t$, $DiffTemp_t$
1-4	$InTemp_{t+1}$, $InHum_{t+1}$ = prediction results of the LSTM from: $Power_t$
	END FOR

where $InTemp_t$ Indoor temperature at time t, $InHum_t$ Indoor humidity at time t, $OutTemp_t$ Outdoor temperature at time t, $TargetTemp_t$ Target temperature at time t, $FanMode_t$ Fan mode of the air conditioner at time t, $Power_t$ Power Consumption of the air conditioner at time t, $DiffTemp_t$ Difference in temperature at time t

3 Implementation Results

3.1 Accuracy of Prediction

To verify the data analytics model, the prediction results of the ANN and the LSTM from the input dataset of a test dataset are compared with the output dataset of the test dataset. For power consumption prediction, the power consumption dataset of the test dataset is compared with the power consumption prediction results from the ANN. The comparison shows that the power consumption prediction results have a coefficient of determination (R^2) of 0.798 [12] and a root mean square error (RMSE) of 99.183 W [13], indicating that the accuracy of the ANN for prediction is satisfactory. An example of a comparison of the power consumption prediction results dataset and the power consumption dataset of the test dataset is shown in Fig. 5.

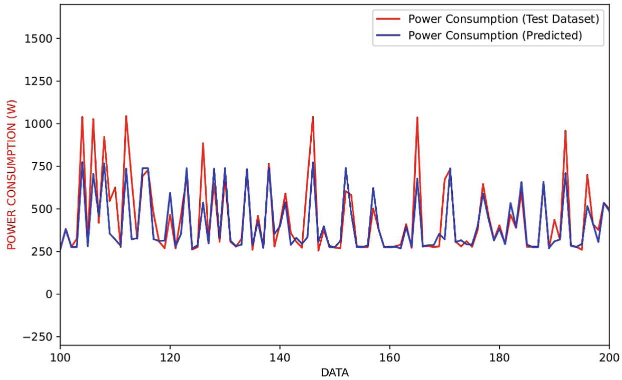


Fig. 5. Comparison of the power consumption from the prediction results by ANN and the output of the test dataset

For indoor temperature and indoor humidity prediction, the results of indoor temperature and indoor humidity obtained from the LSTM are compared with the indoor temperature and indoor humidity data set of the test data set and shown in Fig. 6. The comparison shows that the predicted results for indoor temperature were an R^2 of 0.998 and an RMSE of 0.062 °C, and that the predicted results for indoor humidity were an R^2 of 0.999 and an RMSE of 0.089 rh. It is obvious that the LSTM model produces very accurate predictions because the prediction and the output time series are almost the same.

3.2 Prediction Results

For ambient conditions with an outdoor temperature of 30.25 °C, an indoor temperature of 29.36 °C, and an indoor humidity of 64.66 rh, two scenarios of different temperature and fan mode settings were tested. After the air conditioner was adjusted, the data analytics model by the ANN and the LSTM would be used to predict the power consumption of the air conditioner, indoor temperature, and indoor humidity for the next period (each

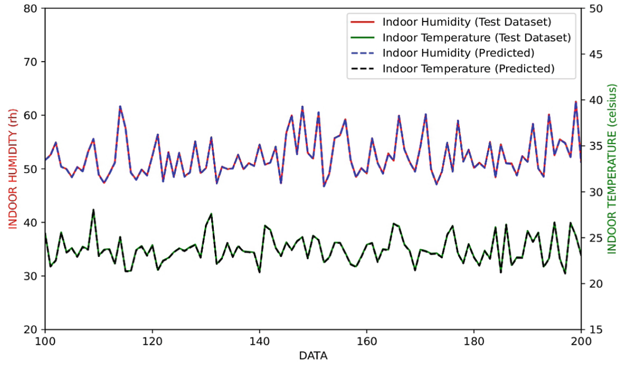


Fig. 6. Comparison of the indoor ambient from the prediction results by LSTM and the output of the test dataset

period was 3 h long). Figure 7 shows an example of the prediction results for the first scenario when the target temperature was 23 °C and the fan mode of the air conditioner was set at 2 (high).

Figure 7 shows that at the beginning, the predicted power consumption was very high, and the predicted indoor temperature and indoor humidity decreased rapidly as well. This behavior was consistent with the operation of the air conditioner, which would accelerate the compressor motor when cooling was required in the first period. Subsequently, when the predicted indoor temperature approached or fell to the target temperature, the predicted power consumption decreased, corresponding to the air conditioning system decelerating the motor when the indoor temperature was slightly colder than the target temperature. In the end, the predicted power consumption, the predicted indoor temperature, and indoor humidity remained constant, allowing the compressor to work at a slower speed to a near-stop state.

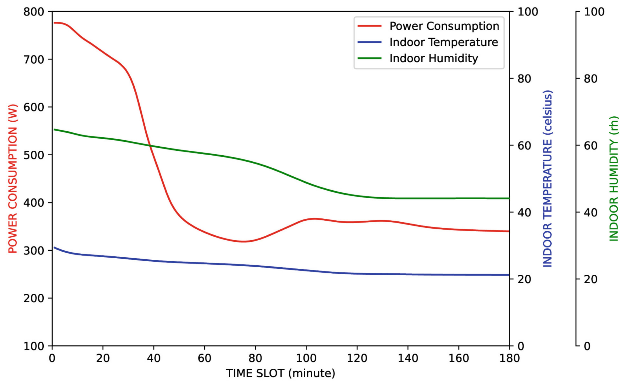


Fig. 7. Example of time-series prediction results for a temperature of 23 °C and fan mode at 2

4 Conclusion

This paper presents a method for predicting the power consumption of an air conditioner and indoor ambient conditions when the air conditioner is adjusted. Which is evaluated from the adjustment of the air conditioner in combination with the effect received from ambient. The data analysis model uses an ANN to predict power consumption and an LSTM to predict indoor temperature and indoor humidity. The experimental results demonstrate that the data analysis model by the ANN and the LSTM can reliably predict the changing behavior of the power consumption of the air conditioner and the ambient conditions in the air-conditioned room from the air-conditioning adjustment and the ambient.

The problems encountered in this research were in the process of collecting data on air conditioning adjustment as the data were collected from the IR Blaster. Often, the residents might adjust the air conditioner directly through the remote control, resulting in discrepancies in the IR blaster that collects the air conditioning adjustment data.

Future research can be extended in the form of a cooling management system for energy savings by utilizing the data analysis model to predict its power consumption and indoor ambient conditions. Therefore, the system can be trained to select the most suitable adjustment of the air conditioner to maximize energy efficiency while keeping the resultant indoor ambient conditions at a level that the occupants are comfortable with.

References

1. Xu, X., Xu, Z., Zhang, B., Guan, X., Jia, Q.-S.: On joint control of heating, ventilation, and air conditioning and natural ventilation in a meeting room for energy saving. Asian J. Control **18**(5), 1781–1804 (2016). https://doi.org/10.1002/asjc.1260
2. Yang, I.-H.: Development of an artificial neural network model to predict the optimal pre-cooling time in office buildings. J. Asian Architect. Build. Eng. **9**(2), 539–546 (2010). https://doi.org/10.3130/jaabe.9.539
3. ISO 7730.: Ergonomics of the thermal environment—analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. International Organization Standardization (2005)
4. Scikit Learn Developers.: Sklearn preprocessing standard scaler. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler (n.d.). Accessed on Jan 2023
5. Keras API Reference.: LSTM layer. Retrieved from https://keras.io/api/layers/recurrent_layers/lstm/ (n.d.). Accessed on Jan 2023
6. Robler, J.: LSTM with Keras. Retrieved from https://medium.com/@jroessl/lstm-with-keras-4c4db55dd349 (2019). Accessed on Jan 2023
7. Scikit Learn Developers.: Sklearn train test split. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split (n.d.). Accessed on Jan 2023
8. Jain, A.K., Mao, J., Mohiuddin, K.M.: Artificial neural networks: a tutorial. Computer **29**(3), 31–44. IEEE Publisher (1996). https://doi.org/10.1109/2.485891
9. Brownlee, J.: Your first deep learning project in python with Keras step by step. Retrieved from https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/ (2022). Accessed on Jan 2023

10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
11. Brownlee, J.: Time series prediction with LSTM recurrent neural networks in python with Keras. Retrieved from https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/ (2022). Accessed on Jan 2023
12. Scikit Learn Developers.: Sklearn metrics R2 score. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score (n.d.). Accessed on Jan 2023
13. Scikit Learn Developers.: Sklearn preprocessing standard scaler. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error (n.d.). Accessed on Jan 2023

A Machine Learning-Based Approach for Sentiment Analysis of Movie Reviews on a Bangladeshi OTT Platform

Hasnur Jahan, Md. Shohel Arman(✉), Afia Hasan, and Sabikun Nahar Bristy

DIU Data Science Lab, Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh

hasnur35-2297@diu.edu.bd, shohelarman.se@gmail.com

Abstract. Sentiment analysis is the examination of feelings and viewpoints in any kind of literature. Opinion mining is another phrase for sentiment analysis. The data's sentiment analysis is quite helpful, to convey the collective, group, or individual viewpoint. This method is employed to ascertain a person's attitude about a specific source. Huge amounts of data are present on social media and other online platforms in the form of tweets, blogs, status, postings, etc. The movie reviews were examined in this research using a variety of methods. On demand of movies on OTT platform several Facebook reviewer pages has been created in Bangladesh. For this work, almost 1000 Bangla reviews were gathered, containing some English word from Facebook. Customer tones were assessed in movie reviews. We use Unigram, Bigram, and Trigram features with a variety of models, including Decision Tree, Random Forest, Multinomial Naive Bayes, K-Neighbors, and Linear Support Vector Machine in n-grams. Random Forest is the most accurate, with 92.35 percent and 90.03 percent accuracy in Unigram and Trigram, respectively. The most accurate model in Bigram is Decision Tree, which has an accuracy of 89.50 percent. This proposed system will help to analysis reviews and give feedback about a movie.

Keywords: Digital Bangladesh · Text mining · Text classification · Movie review · N-grams · Machine learning

1 Introduction

With the introduction of the Web, several sites such as Facebook, Twitter, LinkedIn, and Instagram enable users to share their opinions on a wide range of subjects, from entertainment to education. These platforms have a massive amount of information in the form of tweets, blogs, status updates, posts, etc. In Bangladesh movie reviewer sites are getting more popular day by day because of several OTT platforms, like Hoichoi, Bioscope, and more. People share their thoughts there more through Facebook. Sentiment analysis looks at the text, reviews, and postings that are available online on various platforms to assess the polarity of emotions including happiness, sorrow, grief, hatred, rage, and affection as well as opinions. Opinion mining determines the text's attitude

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 339–349, 2024.
https://doi.org/10.1007/978-3-031-50151-7_32

toward a specific content source. The use of slang terms, misspellings, abbreviations, repeated characters, regional language, and new, emerging emoticons hamper sentiment analysis. The task of determining the suitable sentiment for each word is therefore important. One of the most active study fields is sentiment analysis, which is also extensively researched in data mining. Because opinions are at the heart of most human actions and behaviors, sentiment analysis is used in practically every economic and social domain. The popularity of sentiment analysis is due to its effectiveness. Sentiment analysis may be applied to thousands of documents. It has several uses because it is a productive process that offers good accuracy:

Buying Goods or Services: Making the right decision while buying goods or services is no longer a difficult task. People can quickly assess reviews and opinions of any good or service and compare competing companies by using sentiment analysis.

Product or Service Quality Improvement: Through opinion mining, manufacturers can gather consumer feedback about their goods or services, whether positive or negative, and then improve and raise the bar on their level of quality.

Recommendation Systems: The system can forecast which item should be recommended and which one should not be recommended by assessing and categorizing people's opinions in accordance with their preferences and interests.

Decision-making: The opinions, feelings, and attitudes of others are crucial considerations. Users read comments and reviews of a product before purchasing it, whether it be a book, outfit, or electronic gadget. These reviews have a significant influence on users' decisions.

Marketing research: Sentiment analysis techniques' output can be applied in this field of study. This method can be used to study consumer perceptions of certain goods, services, or new government regulations.

Flame detection: Sentiment analysis makes it simple to monitor newsgroups, blogs, and social media. This method can identify rude, haughty, or overly emotional language in tweets, postings, forums, and blogs on the internet.

The main focus of the study to develop such a system, which will give feedback about a movie automatically by analyzing the movie reviews of the viewers. In that case you can have a idea about a movie, and you don't need to go through the reviews of the movie.

2 Literature Review

Similar to our study, many papers, studies, and research articles have already been published by a variety of authors. These are some work reviews that go along with our work:

In their study [1], Joscha et al. developed and contrasted several methods for using semantic information to enhance sentiment analysis, including Bag of Words models and n-grams. The semantic relationships between sentences or document elements were not taken into account by prior methods. A. Hogenboom et al. [2] researches did not compare the different methodological approaches or offer a strategy for merging disclosure units in the best way possible. Using Rhetoric Structural Theory (RST), which provides a hierarchical representation at the document level, they sought to enhance sentiment analysis. To determine the average sentiment scores from the Rhetoric Structural Theory (RST) tree, they proposed integrating the grid search and weighting. Feature engineering was used to encode the binary data into the random forest, considerably reducing the complexity of the original RST tree. The researchers came to the conclusion that machine learning increased balanced accuracy and provided a high F1 score of 71.9%. Although numerous studies have been conducted in this area (Bhatia, et al., [3]), these do not take into account the numerical (quantitative) data found in the reviews while acknowledging the polarity of sentiment. Also, because of the significant domain dependency of the training data, it cannot be applied to other domains. They concluded that their suggested method of knowledge engineering based on fuzzy sets was significantly more straightforward, effective, and accurate to the tune of over 72% F1.

In this research [1], Amir Hossein Yazdavar et al. offered a fresh interpretation of the sentiment analysis problem incorporating numerical data in drug evaluations. They looked at sentences with quantitative terms to determine if they were opinionated or not, and they also used fuzzy set theory to determine the polarity expressed. Many doctors from different medical centers were interviewed to create a fuzzy knowledge base.

In his study [4], Dhiraj Murthy defined the functions that tweets serve in political campaigns. He emphasized that despite several research and study efforts to determine Twitter's political activity, no effort was made to determine whether these tweets were Predictive or Reactive. He came to the conclusion in his paper that tweets are less predictive and more reactive. He discovered that electoral success was in no way tied to Twitter success and that numerous social media sites were used to boost a candidate's popularity by creating a buzz about them.

In his paper [5], Ahmad Kamal built an opinion mining framework that makes it easier to analyze objectivity or subjectivity, extract features, and summarize reviews, among other things. He classified the subjectiveness and objectivity of reviews using supervised machine learning. He employed several methods, including Naive Bayes, Decision Tree, Multilayer Perceptron, and Bagging. In line with Kamal's paper, he also increased mining performance by avoiding noise and pointless extraction [6].

In this research [7], Humera Shaziya et al. used the WEKA Tool to categorize movie reviews for sentiment analysis. They improved previous work in sentiment categorization, which classifies opinions into those that indicate positive or negative sentiment. They took into account both the positive and negative attitudes that might be expressed in a single review as well as reviews that contain comments from multiple people in this study. They carried out their experiment on WEKA and found that Nave Bayes performs significantly better than SVM for both text and movie reviews. The accuracy of Naive Bayes is 85.1%.

The dataset was constructed using tweets from movie reviews and related tweets about those films by Akshay Amolik et al. in their study [8]. These tweets are subjected to sentiment analysis at the sentence level. There are three stages to it. Preprocessing is the first step. Finally, utilizing pertinent features, a feature vector is produced. Ultimately, tweets were divided into positive, negative, and neutral classes using a variety of classifiers, including Naive Bayes, Support vector machines, Ensemble classifiers, k-means, and Artificial Neural Networks. The findings indicate that we obtain an SVM accuracy of 75%. He refuted the assertion made in the Wu et al. research [9], which noted that the presence of a @username in a tweet influences probability and an action. Nevertheless, Akshay Amolik changed @username in this paper to AT USER, and hashtags were also eliminated as a result. As a result, we employed Support Vector Machine rather than Naive Bayes, which improved accuracy by 10%. Jahan H. and et al. performed a sentiment analysis on customer reviews of e-commerce in Bangladesh. Here they used Naive Bayes, Decision Tree, Multilayer Perceptron, random forest, Linear Support Vector Machine, and k-neighbors used in this work [16].

In previous work there we can see authors mostly have worked on different topics for having a system that will analyze the sentiment of text and will give feedback about that text. This work is also an extension of the previous work.

3 Methodology

Data pre-processing and data collection were our first steps. The parameters for our training method were then optimized. Using our dataset, we then trained and evaluated it. Then, to verify that our algorithms were accurate, we put our output to the test. Lastly, evaluate the correctness of the algorithms. Here used all those methodologies because all those methods are widely used in this type of work. But these works is focused have a better result than the previous work.

4 Data Collection

The data collection part is the most vital part of any work. A good data set provides the best results. In our work, we have collected our data set from Facebook. Based on the movie on the OTT platform there have several movie review pages have grown up. First, we have selected some pages of a movie review. On every page, for different types of movies, we have collected the review. Which is the comment section. People share their thoughts about a movie in the comment section. We have collected those movies using a web scrapping process. Instant data scrapper is a web scrapper which is easy to use. It is used by adding it in the chrome extension part. With this web scrapping tool, we collected comments from the comment section.

5 Dataset Description

By web scraping, we have gathered a total of 1000 reviews from Facebook. When we collected our data through web scrapping there have several parts in the dataset. But for our work, we need 2 features only in the dataset. So, in the final dataset, we keep two features. The first is reviewer comments, while the second is the type of data. We have changed the status of the reviews manually. The reviews which were good about the movie were categorized as “Positive”, then all negative reviews are specifically identified and marked as “Negative”, and lastly the impartial reviews were marked as “Neutral”. The number of good reviews is 359, the number of negative reviews is 390, and 251 neutral reviews are included in the dataset. Every one of our comments was written in Bangla, with some of them also containing English. The dataset was in xlsx format. In (Fig. 1) there have shown a sample of the dataset.

	comment	tag
0	ভাই এই চাবি টি অনেক ভালো লা	good
1	হুট করে রিলিজ দিয়েই বুদ্ধিমানে	good
2	জানোয়ার দেখাচ্ছিলাম ৪জন মিলে	neutral
3	এটা দেখলে তো মনে হচ্ছে প্রতি	neutral
4	কাল্কেই দেখলাম যতটা ভয়ানক	neutral
...	...	...
995	বাংলাদেশের প্রেক্ষাপটে মুভিটা	good
996	আমার যথেষ্ট ভালো লাগছে .. hoi	neutral
997	কাল দেখেছি। ভালো লেগেছে	good
998	বাঙালি মানে কাপি	bad
999	অনেকটা কি Netfilx এর Darlings	bad

Fig. 1. Movie review

6 Data Preprocessing

The most crucial step in applying a machine learning model to a dataset is preparing the acquired data. Preprocessing the data will aid in producing high-quality results. At first, dropped all the null values from the dataset. A null value means if there have any data field which has no data, that will be dropped. To eliminate extraneous punctuation from the collected comment, import the re-library here. The lambda function was then used to eliminate short comments that were no more than two words. Using that procedure, we eliminated122 minor reviews, increasing the total number of comments to 878.

To determine the target value that one determines after viewing an object, a decision tree model is employed. It is a particular type of prediction model utilized in the fields of statistics, data mining, and machine learning in computer science. It is also known as categorization trees. Here, the target variable is typically treated as a set of discrete values. Class labels are represented by tree nodes, and conjunction qualities that uphold the class labels are shown by the branches. If the target variable is continuous, the regression tree is what is used. The decision tree is used to make the decision. It is used to describe the data in data mining. If nodes are chosen at random, issues could arise. There are several strategies used as a result. Akin to entropy. Entropy measures the randomness of the information that the tree is processing. The randomness equation is (1):

$$E(S) = \sum_{i=1}^{C} -P_i \log_2 p_i \quad (1)$$

7 Random Forest

Random forest is a type of ensemble learning used for classification. A random forest set decision trees classifier was used in various subsamples. It is a meta estimator that uses the mean of the subsamples to increase prediction accuracy and reduce overfitting. It typically practiced bagging during training. The process of combining learning models is called bagging. Regression analysis and classifier are both employed. In the model for growing trees, it introduces randomness. From the subset of the random split, the best feature is found. The random subset of a feature is taken into account for the splitting node. Imputation, grouping, and feature selection using random forest are utilized for missing data.

8 Multinomial Naïve Bayes

It typically has applications in natural language processing. A probabilistic learning approach is used. It makes predictions about each tag's probability based on the provided samples. It calculates each tag's probability. The output from the highest probability is then displayed. It is a mixture of some algorithms that share a common algorithmic principle. The presence of one characteristic has no impact on the other features. When there are already B predictors, we may calculate the probability of class A. (2):

$$P(A|B) = P(A) \times \frac{P(B|A)}{P(B)} \tag{2}$$

Here P(B) = B's prior probability

P(A) = A's prior probability.

P(B|A) = incident of predictor B for provided class A probability.

9 K-Neighbors

K nearest neighbors is the abbreviation for KNN. This particular form of guided learning. The K value is the piece of information that we intend to categorize. It determines distances from data points that are unknown given the entirety of the data. Data points are sorted based on the distance value from ascending to descending. We will take the first K numbers from the sorted data points. The class with the most data points will be utilized to determine the unknown data points.

10 Linear Support Vector Machine

A particularly well-liked supervised learning method is a linear support vector machine. In reality, both classification and regression problems need headgear. It is typically utilized for classification problems. Its major objective is to create the best line between data so that fresh data can be accurately categorized. Hyperplane is the name of the best decision's boundary. To create the hyperplane, this SVM finds the highest point. Cases of extreme points are referred to as SV (Support Vector) cases, and SVM (Support Vector Machine) algorithms are used:

$$f(x) = w_t x + b = \sum_{k=0}^{n} w_j y_j + b = 0 \tag{3}$$

Here w is one of M-dimensional vector.

B is scalar, and used for defining the hyperplane (3).

11 Performance Measuring

The calculated models are compared and evaluated based on their prediction inaccuracy. In our line of work, we assess f1 score, memory, accuracy, and precision.

Accuracy is the number of comments which is predicted correctly (4):

$$\text{Accuracy} = \frac{\textit{predictions of correct comment}}{\textit{Predictions of all comments}} \tag{4}$$

The percentage of correctly anticipated favorable remarks among all positive comments is known as precision. Formula is (5):

$$\text{Precision} = \frac{\textit{True positive comments}}{\textit{True Positive comments} + \textit{False Positive comments}} \tag{5}$$

F1 is the precision and recalls and weighted average. Equation is (6):

$$\frac{\textit{True Positive comments}}{\textit{True Positive comments} + \frac{\textit{False Negative comments} \times \textit{False Positive comments}}{2}} \tag{6}$$

Recall is the percentage of all comments combined with just the remarks that were successfully anticipated. Formula is (7):

$$\text{Recall} = \frac{\textit{True positive comments}}{\textit{True Positive comments} + \textit{False Negative comments}} \tag{7}$$

12 Result and Discussion

We were able to get rid of extraneous punctuation by applying the Python function to the comment. Then removed 122 reviews of small length. Then there have 878 records altogether once the data has been cleaned. After cleansing the data, we compile a summary of negative, positive, and neutral data again (Fig. 2). The Negative class contains 320 phrases in total, 3456 words, 645 of which are unique, and the top ten words are displayed. The positive remark class contains 389 sentences overall, 8302words, 1710 distinct phrases, and the top ten most often used words are also displayed. There are 169 total phrases in the neutral remark class, 1897 total words, 839 total unique terms, and the top ten most often used words are also shown. Also displayed a length frequency distribution (Fig. 3).

Finally, using only the positive and negative tag comments, a new dataset is created for model development. There are 320 negative data points and 389 positive data points in the dataset. Label encoding was used for the collection of these data. In this procedure, sentiment labels were converted into a Numpy array for datasets. The class name was then defined again. Then, for encoded labels, the labels were returned together with the condition. The dataset was then separated into testing and training data. We divided our dataset using the holdout method of cross-validation. Where the ratio of training data to test data is 80:20.

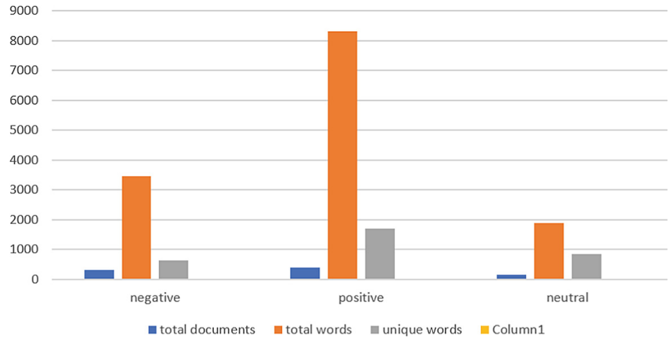


Fig. 2. Data statistics

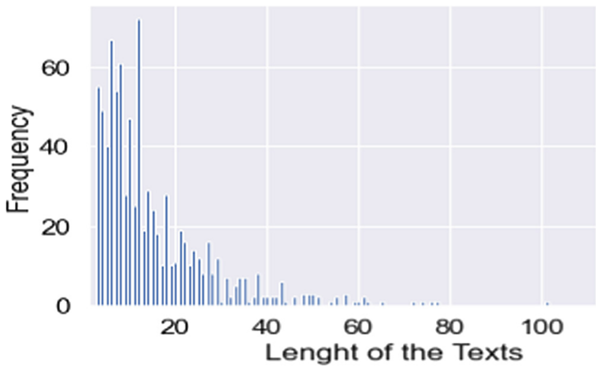


Fig. 3. Length-frequency distribution

Then extracted our feature using the TF-IDF method. And it altered the way we viewed the world. TF-IDF Vectorizer is used to turn every single word of a dataset into numeric form. For preserving the context of vectorized text, n-gram is useful. And Lambda is utilized as a tokenizer there.

On our spitted data, we employed various models. We apply the majority of the classification algorithm because this is a classification problem. We used SVM, SGD Classifier, and Nave Bayes, Decision Tree, Random Forest, Multinomial Naive Bayes, K Neighbors Classifier, and Linear Support Vector machine models to reach optimal results. Without modifying the parameters of our models, we did not modify any models. There used random state of 125 for linear regression, the entropy criterion and 2 random states for the decision tree, 100 estimators and 2 random states for the random forest, 0.10 alpha parameter for multinomial Nave Bayes, 6 neighbors for the K-Neighbors classifier, and 2 random states for both linear and kernel SVM. Each model utilized n-grams features. This application is extensively utilized for natural language processing and text mining. The use of Unigram, Bigram, and Trigram was seen.

We calculated the performance of each model into Unigram, Bigram, and Trigram after fitting our data into the model. There are differences between unigram, bigram, and trigram models. With 92.35% accuracy, Random Forest plays the most important

Table 1. Unigram feature performance

	Accuracy	Precision	Recall	F1 score	Model name
1	83.04	82.59	87.79	83.14	DT
2	92.35	94.64	87.12	91.13	RF
3	85.43	79.19	92.24	83.24	MNB
4	87.05	82.61	91.48	83.36	KNN
5	83.17	94.56	63.25	73.63	Linear SVM

Table 2. Bigram feature performance

1	Accuracy	Precision	Recall	F1 score	Model name
2	89.50	83.35	90.33	85.77	DT
3	87.77	94.23	77.78	85.23	RF
4	85.89	74.25	96.33	83.31	MNB
5	85.05	83.61	92.48	83.36	KNN
6	82.73	95.35	65.08	77.38	Linear SVM

Table 3. Trigram feature performance

	Accuracy	Precision	Recall	F1 score	Model name
1	83.74	82.84	91.42	87.02	DT
2	90.03	100.00	79.78	89.92	RF
3	83.23	72.23	95.21	85.62	MNB
4	85.30	81.43	90.48	85.71	KNN
5	82.53	91.00	60.39	73.34	Linear SVM

function in Unigram feature (Table 1). In Bigram, Decision Tree has the highest accuracy, at 89.50 percent (Table 2). Random Forest had the most accuracy in the Trigram feature, with 90.03 percent accuracy (Table 3).

As we get a higher accuracy with the proposed model this work can be applied to generating a new system that will classify movie reviews more correctly than previous work.

13 Conclusion

Determining the review polarity can be helpful in a variety of areas. Without the user having to read through each review, the user can make decisions based on the results returned by the intelligent systems, which can be developed to give users with comprehensive reviews of movies, products, services, etc.

Nowadays there have several OTT platforms like Hoichoi and a bioscope. And for those, there have created a fanbase for a different type of movies there. Facebook there have several Facebook pages for reviewing those movies. We collected viewers' reviews by scraping these Facebook pages. The information is all in Bengali, while some of it has also been mixed with English information. We use Unigram, Bigram, and Trigram features with a variety of models, including Decision Tree, Random Forest, Multinomial Naive Bayes, K-Neighbors, and Linear Support Vector Machine in n-grams. At 92.35% and 90.03% accuracy in Unigram and Trigram, Random Forest is the most accurate. The Unigram feature and Random Forest are among the best fits. In this essay, we dealt with both favorable and unfavorable customer feedback. The main focus of the work was to create a system that will predict which movie is good and which is bad according to the reviewer's reviews. And a higher accuracy will help to develop a more accurate suggestive system for movie reviews. We intend to deal with impartial reviews and their impact on movie reviews in the future and also aspired to implement the models in a system.

References

1. Erevelles, S., Fukawa, N., Swayne, L.: Big data consumer analytics and the transformation of marketing. J. Bus. Res. **69**(2), 897–904 (2016)
2. Russom, P. et al.: Big data analytics. In: TDWI Best Practices Report, Fourth Quarter, pp. 1–35 (2011)
3. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, pp. 174–181 (1997)
4. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 347–354 (2005)
5. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, p. 271 (2004)
6. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: LREC, vol. 10, no. 2010 (2010)
7. Wahyudi, M., Kristiyanti, D.A.: Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization. J. Theor. Appl. Inf. Technol. **91**(1) (2016)
8. Devi, D.N., Kumar, C.K., Prasad, S.: A feature-based approach for sentiment analysis by using support vector machine. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 3–8. IEEE (2016)

9. Narayanan, V., Arora, I., Bhatia, A.: Fast and accurate sentiment classification using an enhanced naive Bayes model. In: International Conference on Intelligent Data Engineering and Automated Learning, pp. 194–201. Springer (2013)
10. Tan, S., Zhang, J.: An empirical study of sentiment analysis for chinese documents. Expert Syst. Appl. **34**(4), 2622–2629 (2008)
11. Ye, Q., Zhang, Z., Law, R.: Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst. Appl. **36**(3), 6527–6535 (2009)
12. Yu, L.-C., Wu, J.-L., Chang, P.-C., Chu, H.-S.: Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowl.-Based Syst. **41**, 89–97 (2013)
13. Hagenau, M., Liebmann, M., Neumann, D.: Automated news reading: stock price prediction based on financial news using context-capturing features. Decis. Support Syst. **55**(3), 685–697 (2013)
14. Xu, T., Peng, Q., Cheng, Y.: Identifying the semantic orientation of terms using S-HAL for sentiment analysis. Knowl.-Based Syst. **35**, 279–289 (2012)
15. Jahan, H., Kowshir Bitto, A., Shohel Arman, M., Mahmud, I., Fahad Hossain, S., Moni Saha, R., Shohug, M.M.H.: Sentiment analysis of e-commerce consumer based on product delivery time using machine learning. In: Machine Intelligence and Data Science Applications: Proceedings of MIDAS 2021, pp. 649–661. Springer Nature Singapore, Singapore (2022)
16. Maks, I., Vossen, P.: A lexicon model for deep sentiment analysis and opinion mining applications. Decis. Support Syst. **53**(4), 680–688 (2012)
17. Arman, M.S., Islam, A.: Measuring the effectiveness of software code review comments. In: Advances in Computing and Data Sciences: 4th International Conference, ICACDS 2020, Valletta, Malta, Revised Selected Papers, vol. 1244, p. 247. Springer Nature (2020)
18. Islam, F., Arman, M.S., Jahan, N., Sammak, M.H., Tasnim, N., Mahmud, I.: Model and popularity based recommendation system-A collaborative filtering approach. In: 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–5. IEEE (2022)
19. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. J. Mach. Learn. Res. **2**(Nov), 45–66 (2001)
20. Siddikk, A.B., Lia, R.J., Muntasir, M.F., Rahman, S.S.M.M., Arman, M.S., Jahan, M.R.: FakeTouch: machine learning based framework for detecting fake news. In: Baddi, Y., Gahi, Y., Maleh, Y., Alazab, M., Tawalbeh, L. (eds.) Big Data Intelligence for Smart Applications. SCI, vol. 994, pp. 317–334. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-87954-9_15
21. Kumar, S., Yadava, M., Roy, P.P.: Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. Inf. Fusion **52**, 41–52 (2019)
22. Arman, M.S., Alam, M.R., Jahan, H., Islam, L., Sammak, M.H., Biplob, K.B.M.B.: A data mining approach to finding face mask from Bangladeshi news channel. In: 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7. IEEE (2022)
23. Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.): Intelligent computing & optimization. In: Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022), vol. 569. Springer Nature (2022)
24. Vasant, P., Zelinka, I., Weber, G.W.: Intelligent computing & optimization. In: Conference Proceedings ICO, p. 804 (2018)

Pareto Optimization of Laminated Composite Plates with Non-dominated Sorting Whale Optimization Algorithm (NSWOA) Ensemble

G. Shanmugasundar[1], Subham Pal[2], Jasgurpeet Singh Chohan[3], and Kanak Kalita[4](✉)

[1] Department of Mechanical Engineering, Sri Sairam Institute of Technology, Chennai, India
[2] Department of Aerospace Engineering and Applied Mechanics, Indian Institute of Engineering Science and Technology, Shibpur, India
[3] Department of Mechanical Engineering, University Centre for Research & Development, Chandigarh University, Mohali, India
[4] Department of Mechanical Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Avadi, India
drkanakkalita@veltech.edu.in

Abstract. Plates are non-curve structural elements used in civil, aerospace and marine constructions. Optimization of these structures is important for its longevity and safe design. This study uses a novel non-dominated sorting whale optimization algorithm (NSWOA) ensemble technique to optimize the design of laminated composite plates. The objective of the study is to maximize the fundamental frequency and frequency separation. The orientations of the fiber in each layer are considered the design variables. This study considers a rectangular plate with a symmetric arrangement of lamina under simply supported boundary conditions. The results obtained through the implementation of the NSWOA ensemble technique are presented, and this will assist in future research.

Keywords: Pareto front · Optimal design · Composite · Multi-objective · Whale optimizer

1 Introduction

Plates are structural elements mostly used in aerospace, civil, mechanical and marine industries [1]. Due to the advancement in material science, laminated composite plates are being used widely nowadays. Composite materials show very good durability with lightweight, so it is more advantageous than the conventional engineering materials in lightweight applications like aircraft and watercraft casings. The dynamic behaviour of plates in various boundary conditions is essential for the design of safe and efficient structures. The composite plate's free vibration characteristics largely depends on the fibres' orientation within the laminate. So, the optimal design of such fiber angle orientation is essential to maximize the frequency and frequency separation.

Over the last few decades, many researchers proposed different optimization techniques to maximize the fundamental frequency considering fiber angle orientation as

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 350–357, 2024.
https://doi.org/10.1007/978-3-031-50151-7_33

the design variable. In published investigations, the maximization of frequency separation issues has received minimal attention. First investigated by Bronowicki et al. [2] in 1975, frequency separation for ring-stiffened cylindrical shells. The cylindrical shell was subjected to an external pressure load after being reinforced with a T-ring stiffener. Frequency separation was analyzed using an orthotropic shell model, and sequential unconstrained optimization techniques were used for optimization. They intended to reduce the weight of the structure and increase the frequency separation between the first two Eigen-frequencies. They discovered that a moderate increase in weight beyond the optimal value doubles the frequency separation. Pappas [3] investigated the optimal frequency separation for a similar type of shell structure as Bronowicki et al. [2] used. However, it was determined that the carapace was subject to hydrostatic pressure. According to Pappas [3], the gradient-based mathematical programming-based optimization technique employed by Bronowicki et al. [2] was inappropriate for such problems. Pappas [3], therefore, proposed the direct search scheme with a gradient-based direction-finding (DSFD) procedure, which combines a direct search scheme with a gradient-based direction-finding scheme. Adali [4] optimized anti-symmetrically laminated angle-ply plates for the maximization of first natural frequency and frequency separation. The orientation and thickness of each lamina were taken into account during the design process. Duffy and Adali [5] investigated the optimal fiber angle orientation of hybrid laminate for maximizing the first natural frequency and the distance between the first two natural frequencies. They believed that the ply angles were mass-restricted and arranged anti-symmetrically. Du and Olhoff [6] investigated the maximal values of simple and multiple Eigen frequencies as well as frequency gaps of freely vibrating continuum structures. They used topology optimization to determine the beam structure's maximal frequency and frequency separation. Farshi and Rabiei [7] examined the optimal composite laminate design for fundamental frequency and frequency separation. The optimal fibre orientation and thickness were determined using the layerwise optimization algorithm (LOA). Over the years, numerous metaheuristic algorithms have been utilized to optimize the composite laminate. Genetic algorithm (GA) [8], particle swarm optimization (PSO) [9], cuckoo search (CS) [10], LOA [11], artificial bee colony (ABC) [12], differential evolution [13], ant colony optimization (ACO) [14, 15], etc. are frequently employed in structural optimization. GA is one of the earliest and most widely used search strategies. Using branched multipoint approximate functions and GA, An et al. [16] analyzed the multi-objective optimal design of hybrid composite laminate. They aimed to maximize the fundamental frequency and frequency separation while minimizing costs. As design parameters, fiber angle orientations were considered. Using GA, PSO, and CS, Kalita et al. [9] examined multi-objective optimization of skew composite laminates. They attempted to optimize the frequency disparity between the first two natural frequencies and the first natural frequency.

Literature study shows that frequency maximization is very popular among researchers. However, single-objective optimization is common, and optimizers used are outdated. Multi-objective optimization considering the newly proposed optimization technique, is very rare. So, in this article, a novel non-dominated sorting whale optimization algorithm (NSWOA) ensemble optimizer is used to find the optimal solutions for frequency and frequency separation. Frequency and frequency separation are two objective functions in multi-objective optimization, and fiber angle orientation is the design variable.

2 NSWOA

The Non-Dominated Sorting Whale Optimization Algorithm (NSWOA) is an advanced optimization technique that combines the strengths of the Whale Optimization Algorithm (WOA) with non-dominated sorting (NDS) for multi-objective optimization problems. The WOA is a metaheuristic algorithm inspired by the hunting behaviour of humpback whales, while NDS is a technique used for ranking solutions in multi-objective optimization based on their dominance. The NSWOA integrates these approaches to efficiently search the solution space and find a set of optimal solutions (Pareto front). The pseudocode for NSWOA is detailed below:

Algorithm 1: NSWOA

Initialize the population of whales
Evaluate the fitness of each whale
Initialize the archive of non-dominated solutions
Calculate the initial positions of the whales
While stopping criteria not met:
Update the position of each whale using WOA equations
Evaluate the fitness of each whale
Update the archive with the new non-dominated solutions
Perform non-dominated sorting on the combined population (whales + archive)
Update the population of whales based on the sorted solutions
Return the final archive of non-dominated solutions

3 NSWOA Ensemble

The NSWOA ensemble is a novel approach that combines the strengths of multiple Pareto fronts of NSWOA. The NSWOA ensemble is inspired by ensemble machine learning models [17]. It aims to combine the Pareto optimal solutions obtained in multiple independent runs of the NSWOA and combines them into finding an ensemble Pareto front. This ensemble Pareto front is expected to be more effective and robust than its individual counterparts. The pseudocode for the NSWOA ensemble is detailed below:

Algorithm 2: NSWOA ensemble

Set the number of independent trials (n)
For each trial in n trials:
Perform NSWOA optimization and store the Pareto optimal front
Combine the n Pareto optimal fronts from NSWOA trials and apply secondary NDS
Return the final NSWOA ensemble Pareto optimal front

4 Problem Description

The four fiber angles ($\theta_1, \theta_2, \theta_3$ and θ_4) of an 8-layer symmetric all-sides simply-supported composite plate are chosen as the design variables. The upper and lower bounds of the design variable is $[-90, 90]$. The material properties are chosen as $E_1/E_2 = 40, G_{12}/E_2 = 0.6, G_{23}/E_2 = 0.5$ and $\vartheta_{12} = 0.25$. . The composite plate is of rectangular in shape with an aspect ratio of 2 (i.e. $b/a = 2, h/a = 0.1, \alpha = 0,$ and $n = 8$. . The objective is to simultaneously maximize the fundamental frequency (λ_1) and frequency separation between the first two modes (λ_{21}). The λ_1 and λ_{21} for this problem are expressed as functions of the design variables as per the following Eq. (1) and (2) [18]:

$$
\begin{aligned}
\lambda_1 = {} & 64.7921 - 0.00097\theta_1 - 0.01303\theta_2 + 0.00106\theta_3 \\
& + 0.00968\theta_4 + 0.00013\theta_2\theta_3 - 0.00019\theta_3\theta_4 \\
& - 0.00287\theta_1^2 - 0.0017\theta_2^2 - 0.000854\theta_3^2
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\lambda_{21} = {} & 12.31464 - 0.00305\theta_1 - 0.013577\theta_2 - 0.018836\theta_3 \\
& - 0.026293\theta_4 + 0.000417\theta_1\theta_3 + 0.000237\theta_1\theta_4 \\
& - 0.00022\theta_2\theta_3 - 0.000229\theta_2\theta_4 + 0.00165\theta_1^2 \\
& + 0.00075\theta_2^2 - 0.00035\theta_3^2
\end{aligned} \tag{2}
$$

5 Results and Discussion

The multi-objective optimization with fundamental frequency and frequency separation (i.e. frequency gap between first two eigen frequencies) as two objective functions is performed using NSWOA technique. Figure 1 depicts the Pareto front and box plot of five independent trials derived using NSWOA.

The fundamental frequency is plotted along the x-axis, while frequency separation is plotted along the y-axis. The frequency value and frequency separation value obtained from five trials of NSWOA is comparable, as shown in Fig. 1. The mean and standard devastation for fundamental frequency and frequency separation are comparable in each experiment. Thus, the NSWOA can be said to be highly reliable in generating similar optimized results over multiple independent trials.

After obtaining non-dominant Pareto front solutions from the NSWOA, the ensemble method is used to collate and condense the results further. Figure 2 demonstrates that the NSWOA ensemble technique yields an improved Pareto front. This Pareto front contains the NDS solutions from the all previous five NSWOA trials. Figure 2 depicts the Pareto front solutions derived from the NSWOA ensemble technique as well as the box plot of all functional evaluations. Each point in Fig. 2 represents an optimal solution, allowing someone to design a plate based on the design requirements.

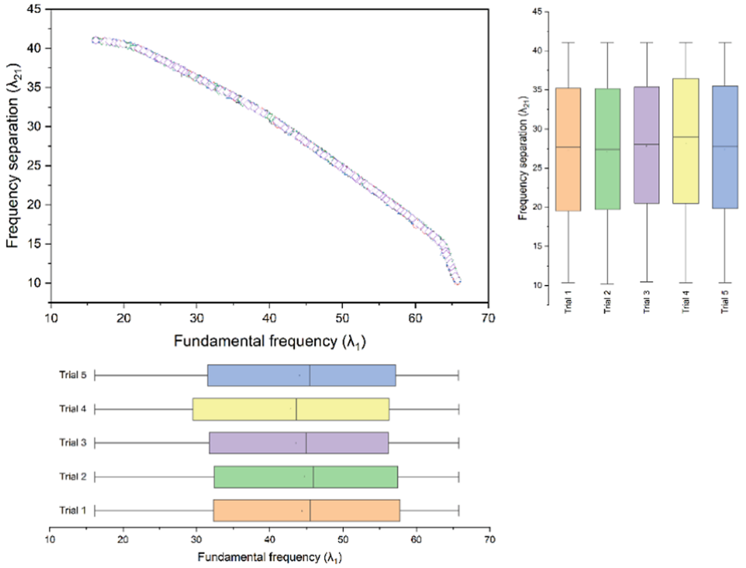


Fig. 1. Pareto fronts of 5 independent trials of NSWOA along with box plots

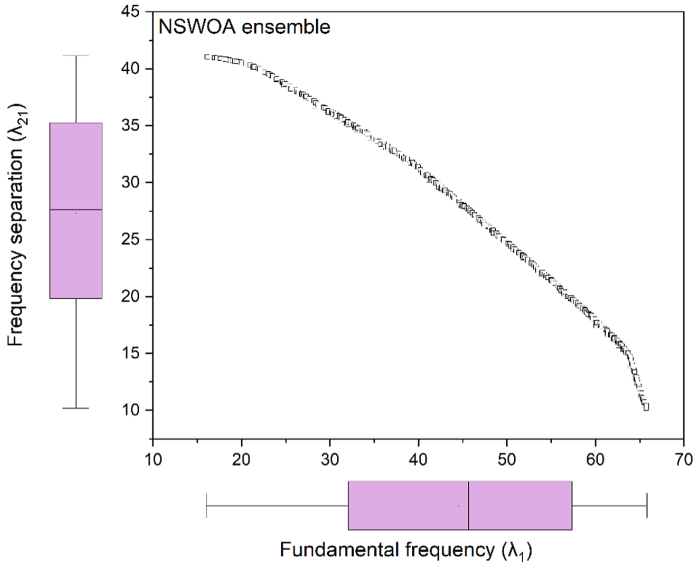


Fig. 2. Pareto front of NSWOA ensemble along with box plots

Figure 3 demonstrates the relationship between the optimal fiber angle orientation and the fundamental frequency and frequency separation derived from the NSWOA ensemble. This interaction between optimized design variables and responses will assist a decision-maker in selecting the optimal design variable based on the required response.

Figures 4 and 5 shows the mode shape of first three eigen frequencies of the two extremum points of the optimized NSWOA ensemble Pareto front. The two extremum points of the NSWOA ensemble Pareto fronts are at the following combinations:

$$\theta_1 = -90, \theta_2 = 90, \theta_3 = -90, \theta_4 = -90, \lambda_1 = 16.14, \lambda_{21} = 41.04;$$
$$\theta_1 = 0.02544, \theta_2 = -7.4117, \theta_3 = -11.17672, \theta_4 = 90, \lambda_1 = 65.75, \lambda_{21} = 10.39;$$

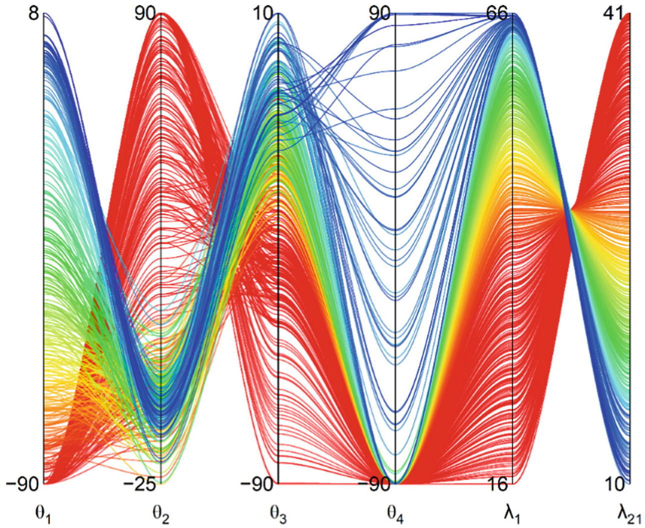


Fig. 3. Interaction between optimized design variables and responses (NSWOA ensemble)

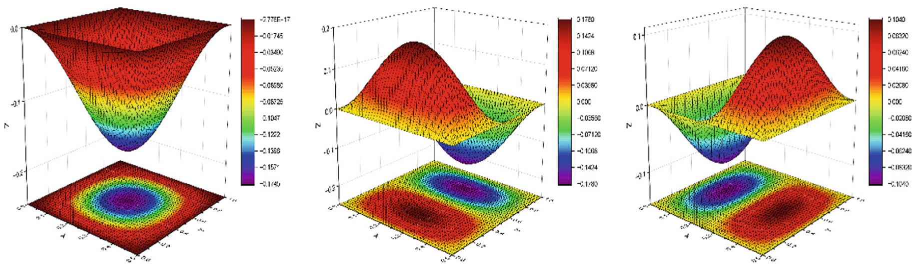


Fig. 4. Mode shape of first three eigen frequencies of symmetric laminate for fiber angle orientation of $[-90/90/-90/-90]_s$

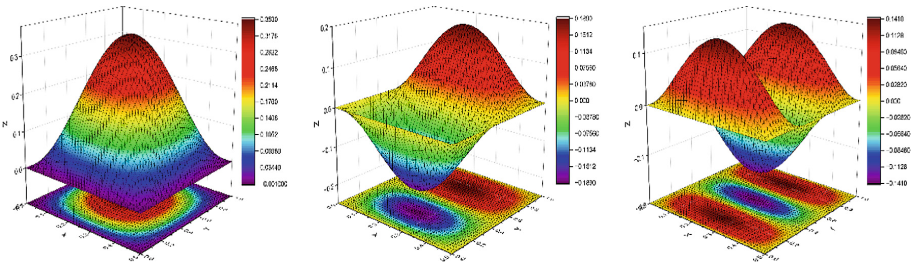


Fig. 5. Mode shape of first three eigen frequencies of symmetric laminate for fiber angle orientation of $[0.025438/-7.41169/-11.1762/90]_s$

6 Conclusion

This study proposes a novel NSWOA ensemble technique to determine the optimal design of composite laminated plates. This research aims to maximize the fundamental frequency and frequency separation value using fiber angle orientations as design variables. A laminated rectangular plate with simply supported boundary conditions is analyzed to determine the response from the chosen design elements. A second-order polynomial model is developed based on the responses to establish a relationship between fiber angle orientation and response parameters (i.e., frequency and frequency separation). In NSWOA optimization techniques, these models serve as the objective function. Using the NSWOA ensemble, the Pareto front derived with NSWOA is further shortened by selecting the non-dominant solutions. It is observed that the novel NSWOA consistently generates identical outcomes. The presented results will be of great assistance in designing a plate to meet the design requirements.

References

1. Pal, S., Kalita, K., Haldar, S.: Dynamic analysis of laminated shell panels with cut-out and cracked corners carrying concentrated and distributed mass. Mech. Based Des. Struct. Mach. (2023)
2. Bronowicki, A.J., Nelson, R.B., Felton, L.P., Schmit, L.A., Jr.: Optimization of ring stiffened cylindrical shells. AIAA J. **13**, 1319–1325 (1975)
3. Pappas, M.: Optimal frequency separation of cylindrical shells. AIAA J. **16**, 999–1001 (1978)
4. Adali, S.: Design of shear-deformable antisymmetric angle-ply laminates to maximize the fundamental frequency and frequency separation. Compos. Struct. **2**, 349–369 (1984)
5. Duffy, K.J., Adali, S.: Optimal fibre orientation of antisymmetric hybrid laminates for maximum fundamental frequency and frequency separation. J. Sound Vib. **146**, 181–190 (1991)
6. Du, J., Olhoff, N.: Topological design of freely vibrating continuum structures for maximum values of simple and multiple eigenfrequencies and frequency gaps. Struct. Multidiscip. Optim. **34**, 91–110 (2007)
7. Farshi, B., Rabiei, R.: Optimum design of composite laminates for frequency constraints. Compos. Struct. **81**, 587–597 (2007)
8. Dey, S., Mukhopadhyay, T., Khodaparast, H.H., Adhikari, S.: A response surface modelling approach for resonance driven reliability based optimization of composite shells. Periodica Polytech. Civ. Eng. **60**, 103–111 (2016)

9. Kalita, K., Ragavendran, U., Ramachandran, M., Bhoi, A.K.: Weighted sum multi-objective optimization of skew composite laminates. Struct. Eng. Mech. **69**, 21–31 (2019)
10. Kalita, K., Dey, P., Haldar, S., Gao, X.-Z.: Optimizing frequencies of skew composite laminates with metaheuristic algorithms. Eng. Comput. **36**, 741–761 (2020)
11. Narita, Y., Robinson, P.: Maximizing the fundamental frequency of laminated cylindrical panels using layerwise optimization. Int. J. Mech. Sci. **48**, 1516–1524 (2006)
12. Apalak, M.K., Karaboga, D., Akay, B.: The artificial bee colony algorithm in layer optimization for the maximum fundamental frequency of symmetrical laminated composite plates. Eng. Optim. **46**, 420–437 (2014)
13. Vo-Duy, T., Ho-Huu, V., Do-Thi, T.D., Dang-Trung, H., Nguyen-Thoi, T.: A global numerical approach for lightweight design optimization of laminated composite plates subjected to frequency constraints. Compos. Struct. **159**, 646–655 (2017)
14. Abachizadeh, M., Tahani, M.: An ant colony optimization approach to multi-objective optimal design of symmetric hybrid laminates for maximum fundamental frequency and minimum cost. Struct. Multidiscip. Optim. **37**, 367–376 (2009)
15. Koide, R.M., Luersen, M.A.: Ant colony algorithm applied to fundamental frequency maximization of laminated composite cylindrical shells. In: Brazilian Conference on Composite Materials (2012)
16. An, H., Chen, S., Huang, H.: Multi-objective optimal design of hybrid composite laminates for minimum cost and maximum fundamental frequency and frequency gaps. Compos. Struct. **209**, 268–276 (2019)
17. Kalita, K., Kumar, V., Chakraborty, S.: A novel MOALO-MODA ensemble approach for multi-objective optimization of machining parameters for metal matrix composites. Multiscale Multidisc. Model. Exp. Des. 1–19 (2023)
18. Kalita, K., Dey, P., Joshi, M., Haldar, S.: A response surface modelling approach for multi-objective optimization of composite plates. Steel Compos. Struct. Int. J. **32**(4), 455–466 (2019)

A Survey on Smart Intelligent Computing and Its Applications

Ramsagar Yadav[1](✉), Mukhdeep Singh Manshahia[1], Mahendra Pal Chaudhary[2], and Doopinder Kaur[3]

[1] Department of Mathematics, Punjabi University, Patiala, Punjab, India
ramsagar.yadav@lsraheja.org
[2] International Scientific Research and Welfare Organization, New Delhi, India
[3] Government Mohindra College, Patiala, India

Abstract. Smart intelligent computing has emerged as a key technology in the field of computer science and engineering. It has revolutionized the way we interact with machines and has paved the way for the development of advanced applications in various domains. This survey paper presents an overview of smart intelligent computing and its various applications. The paper discusses the basic concepts, architecture, and components of smart intelligent computing. It then goes on to discuss the various applications of smart intelligent computing in areas such as healthcare, transportation, manufacturing, and energy management. The paper also examines some of the challenges faced by the technology and its potential future developments.

Keywords: Smart computing · Intelligent computing · Machine learning · Artificial intelligence · Applications · Challenges

1 Introduction

In recent years, there has been a significant increase in the development and implementation of smart intelligent computing systems. These systems utilize cutting-edge technologies such as artificial intelligence, machine learning, and deep learning to enable intelligent decision-making and automation. The potential applications of smart intelligent computing are vast and include areas such as healthcare, transportation, finance, and education, among others [1]. This paper aims to provide a comprehensive survey of smart intelligent computing and its applications. Smart intelligent computing refers to the use of advanced technologies such as artificial intelligence (AI), machine learning (ML), and deep learning (DL) to develop computer systems that can learn from data and make intelligent decisions [2]. These technologies allow computers to learn and improve their performance over time without being explicitly programmed. Smart intelligent computing systems can process large amounts of data quickly and accurately, allowing them to make informed decisions in real-time [3].

One of the key advantages of smart intelligent computing is its ability to automate repetitive and time-consuming tasks. By automating these tasks, businesses can improve

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 358–365, 2024.
https://doi.org/10.1007/978-3-031-50151-7_34

efficiency and productivity, reduce errors, and free up employees to focus on more complex and strategic work [4]. Another important application of smart intelligent computing is in the healthcare industry. Intelligent systems can analyze large amounts of patient data to identify patterns and make accurate diagnoses. They can also help healthcare providers develop personalized treatment plans for patients based on their individual health profiles [5]. Transportation is another area where smart intelligent computing is being used to improve safety and efficiency. For example, self-driving cars are being developed that can use AI and ML algorithms to make decisions in real-time and navigate traffic safely [6].

Smart intelligent computing is also being used in finance to analyze large amounts of financial data and identify patterns and trends. This information can be used to make more informed investment decisions, detect fraudulent activity, and manage risk [7]. In the field of education, smart intelligent computing is being used to develop personalized learning experiences for students. By analyzing data on each student's learning style and progress, intelligent systems can create customized learning plans that are tailored to their individual needs and abilities [8].

Smart intelligent computing is also playing a crucial role in the development of smart cities. Intelligent systems can monitor traffic flow, manage energy consumption, and optimize waste management, leading to more sustainable and efficient cities [9]. Overall, smart intelligent computing is transforming the way we live and work, improving efficiency, accuracy, and decision-making across a wide range of industries and applications. However, it is important to ensure that these systems are designed and implemented ethically, with a focus on transparency, accountability, and fairness.

2 Organization of Paper

The paper is organized into several sections. First, we provide an introduction to the topic and highlight the motivation for this research. Next, we conduct a literature review to explore the current state of the art in smart intelligent computing. We then conduct a gap analysis to identify areas where further research is needed. Following this, we present a comparative analysis of various smart intelligent computing techniques and their applications in tabular form. We then present our findings and suggestions for future research in this area. We also discuss the advantages and limitations of our research. Finally, we conclude the paper with a summary of our key findings and suggestions for future research.

3 Motivation of Research

The motivation for this research stems from the increasing importance of smart intelligent computing in various fields. The potential benefits of these systems are numerous, including increased efficiency, accuracy, and cost-effectiveness. However, there is still much to be learned about the applications of smart intelligent computing and the techniques used to implement them. This research aims to provide a comprehensive overview of the current state of the art in this field and identify areas where further research is needed.

Additionally, there are ethical and privacy concerns that must be addressed with the increasing adoption of smart intelligent computing systems. As these systems become more integrated into our daily lives, it is essential to ensure that they are developed and implemented in a responsible and ethical manner. Therefore, this research also aims to examine the ethical implications of smart intelligent computing and identify ways to mitigate any potential negative consequences.

The motivation for this research is to provide a comprehensive understanding of the current state of smart intelligent computing and its applications, as well as to identify areas where further research is needed to ensure its responsible and ethical implementation.

4 Literature Review

The literature review explores the current state of the art in smart intelligent computing. We examine various techniques used in smart intelligent computing, including machine learning, deep learning, and artificial intelligence. We also explore the various applications of these techniques, including in the areas of healthcare, transportation, finance, and education.

Machine learning is a widely used technique in smart intelligent computing. It involves training computer algorithms to learn patterns and make predictions based on data. Various machine learning algorithms have been developed, including decision trees, random forests, support vector machines, and neural networks. Deep learning, a subset of machine learning, involves the use of deep neural networks to learn complex patterns from data. This technique has shown remarkable success in areas such as image recognition, natural language processing, and speech recognition [10].

Artificial intelligence (AI) involves the development of intelligent machines that can perform tasks that typically require human intelligence. This includes tasks such as perception, reasoning, and decision-making. AI techniques are widely used in smart intelligent computing, including in areas such as robotics, natural language processing, and autonomous systems [11]. In the healthcare domain, smart intelligent computing is being used for various tasks such as disease diagnosis, drug discovery, and patient monitoring. For example, machine learning algorithms are being used to predict the risk of developing certain diseases based on patient data [12].

In the transportation domain, smart intelligent computing is being used for tasks such as traffic management, route optimization, and autonomous driving. For example, deep learning algorithms are being used to recognize objects on the road and make decisions in real-time to avoid accidents [13]. In the finance domain, smart intelligent computing is being used for tasks such as fraud detection, risk management, and portfolio optimization. For example, AI techniques are being used to identify fraudulent transactions and prevent financial losses [14].

In the education domain, smart intelligent computing is being used for tasks such as personalized learning, student performance prediction, and plagiarism detection. For example, machine learning algorithms are being used to personalize the learning experience for individual students based on their learning style and performance [15].

5 Gap Analysis

The gap analysis identifies areas where further research is needed in the field of smart intelligent computing. These include the development of more robust and accurate machine learning algorithms, the integration of smart intelligent computing into existing systems, and the ethical and privacy implications of these systems. Additionally, there is a need for more research on the optimization of resource allocation and energy efficiency in smart intelligent computing systems, as well as the development of standardized protocols and frameworks for interoperability between different smart intelligent computing systems.

Furthermore, there is a gap in research on the integration of smart intelligent computing with emerging technologies such as blockchain, edge computing, and the Internet of Things (IoT) to create more advanced and secure systems. Another area for further research is the development of smart intelligent computing systems that can operate in dynamic and uncertain environments, such as disaster response or military operations.

Finally, there is a need for more research on the social and economic impacts of smart intelligent computing, including its potential to displace human workers and the ethical considerations around decision-making by intelligent systems. Overall, there is a significant need for further research in the field of smart intelligent computing to fully realize its potential and address its challenges.

6 Comparative Analysis

The comparative analysis presents a comparison of various smart intelligent computing techniques and their applications in tabular form. We examine the strengths and limitations of each technique and provide examples of their applications in various fields (Table 1) [1–5].

In summary, each technique has its strengths and limitations, and its application depends on the specific needs and goals of the organization or industry. AI, ML, and DL are useful for processing and analyzing large amounts of data quickly and accurately, while cognitive computing is ideal for natural language processing and understanding unstructured data. Robotics are useful for performing repetitive or dangerous tasks, but they can be expensive to design and build. Ultimately, the choice of which technique to use depends on the specific problem and the resources available.

7 Findings and Suggestions

Our findings indicate that smart intelligent computing has the potential to revolutionize various industries, including healthcare, transportation, finance, and education. However, there is still much to be learned about the best practices for implementing these systems and the ethical implications of their use. We suggest further research into the development of more accurate and robust machine learning algorithms, the integration of smart intelligent computing into existing systems, and the ethical considerations of these systems. Based on the analysis, we have found that smart intelligent computing techniques have a wide range of applications in various industries and have the potential

Table 1. Ccomparison of various smart intelligent computing techniques

Technique	Description	Strengths	Limitations	Applications
Artificial intelligence (AI)	The ability of machines to perform tasks that typically require human intelligence, such as learning, problem-solving, and decision-making	Can learn from vast amounts of data and adapt to new situations. Can process information faster than humans	Lack of human judgment and intuition. Can be biased based on the data used to train the model	Healthcare, finance, transportation, education, customer service, and more
Machine learning (ML)	A subset of AI that uses algorithms to learn patterns in data and make predictions or decisions without being explicitly programmed	Can process and analyze large amounts of data quickly and accurately. Can improve over time with more data	Requires a lot of data for effective training. Can be biased based on the data used to train the model	Image and speech recognition, fraud detection, recommendation systems, and more
Deep learning (DL)	A subset of ML that uses neural networks to simulate human brain function and perform complex tasks	Can learn and process complex data and patterns. Can achieve higher accuracy than traditional ML algorithms	Requires a lot of computing power and resources. Can be difficult to interpret how the model arrives at its decision	Natural language processing, image and speech recognition, self-driving cars, and more
Cognitive computing	A system that can learn, reason, and understand natural language like a human	Can understand and process natural language and unstructured data. Can improve over time with more data	Can be expensive to implement and maintain. Requires a lot of computing power and resources	Healthcare, finance, customer service, and more
Robotics	The use of machines to perform tasks that are typically done by humans	Can work in dangerous or inaccessible environments. Can perform repetitive tasks without getting tired	Limited to the tasks it is programmed to do. Can be expensive to design and build	Manufacturing, healthcare, transportation, and more

to transform the way we work and live. However, there are also limitations and challenges associated with these techniques, such as biases and the need for large amounts of data and computing power.

To further advance the field of smart intelligent computing, we suggest several areas of further research. Firstly, there is a need for the development of more accurate and robust machine learning algorithms, particularly in the area of deep learning, which can improve the accuracy and interpretability of these models. Secondly, the integration of smart intelligent computing into existing systems needs to be explored further, particularly in areas such as healthcare and finance, where the integration of these systems can have a significant impact on patient outcomes and financial decision-making. Finally, there is a need for further research into the ethical implications of smart intelligent computing, particularly in the areas of privacy, bias, and accountability.

8 Advantages and Limitations of Research

The advantages of this research include providing a comprehensive overview of the current state of the art in smart intelligent computing and identifying areas where further research is needed. However, the limitations of this research include the limited scope of the literature review and the subjective nature of the comparative analysis. The advantages of this research are that it provides a comprehensive overview of the current state of smart intelligent computing and its applications. The research identifies the strengths and limitations of different techniques and suggests areas for further research, which can be helpful for practitioners and researchers in the field.

However, this research also has some limitations. One of the limitations is the limited scope of the literature review. The research is based on the knowledge cutoff of 2021, and new developments in the field may not have been included. Additionally, the literature review may not have covered all relevant research articles in the field, which may have limited the scope of the analysis. Another limitation is the subjective nature of the comparative analysis. The analysis is based on the author's interpretation of the strengths and limitations of each technique and may not reflect the opinions of others in the field. Additionally, the analysis may not consider the specific context in which these techniques are used, which may limit the generalizability of the findings.

Despite these limitations, this research provides valuable insights into the field of smart intelligent computing and identifies areas where further research is needed. The findings can be used to guide future research and development in this area, ultimately leading to the development of more accurate and effective intelligent computing systems.

9 Conclusion and Future Scope

In conclusion, smart intelligent computing has the potential to revolutionize various industries, but there is still much to be learned about the best practices for implementing these systems and the ethical implications of their use. Future research should focus on the development of more accurate and robust machine learning algorithms, the integration of smart intelligent computing into existing systems, and the ethical considerations of these systems.

The future scope of research in smart intelligent computing is vast, and there are many areas that require further exploration. One area of future research is the development of more accurate and efficient algorithms for smart intelligent computing. This can involve exploring new techniques for machine learning, such as deep learning or reinforcement learning, to improve the accuracy and interpretability of these models. Another area of future research is the integration of smart intelligent computing into existing systems. This can involve exploring how these systems can be used to improve decision-making processes in industries such as healthcare, finance, and transportation. Additionally, the development of new applications and use cases for smart intelligent computing can be explored. Finally, there is a need for further research into the ethical considerations of smart intelligent computing. This can involve exploring how biases in these systems can be identified and addressed, as well as considering the impact of these systems on privacy and security.

Acknowledgements. Authors are grateful to Punjabi University, Patiala for providing adequate library and internet facility.

References

1. Chen, C., Wang, C.: Intelligent sensing with deep learning for internet of things applications: a survey. IEEE Internet Things J. **8**(7), 5466–5485 (2021). https://doi.org/10.1109/JIOT.2021.3091115
2. Zhang, X., Xu, J., Xu, C.: An intelligent sensing system based on deep learning for non-destructive detection of fruit freshness. Comput. Electron. Agric. **182**, 106033 (2021). https://doi.org/10.1016/j.compag.2021.106033
3. Li, X., Li, J., Wang, H.: Recent developments in intelligent sensing and signal processing for industrial systems. Sensors **20**(23), 7018 (2020). https://doi.org/10.3390/s20237018
4. Hajian-Tilaki, K., Hosseini, M.J.: The significance of machine learning in healthcare: features, pillars and applications. J. Med. Syst. **45**(9), 102 (2021). https://doi.org/10.1007/s10916-021-01766-8
5. Raza, S.A., Abid, S., Raza, S.M.: Machine learning techniques and applications in healthcare: a review. Healthc. Technol. Lett. **8**(6), 249–255 (2021). https://doi.org/10.1049/htl2.12012
6. Song, M., Guo, Q.: Efficient 3D object recognition in mobile edge environment. J. Cloud Comput. **11**(1) (2022)
7. Yu, L., Zhang, L., Gong, Z.: An optimization model for landscape planning and environmental design of smart cities based on big data analysis. Sci. Program. **2022**, 1 (2022)
8. Liang, K., Wan, Y., Jiang, Y.: The effect of jogging on the physical health of college students based on intelligent computing. Mob. Inf. Syst. **2022**, 1 (2022)
9. Zhang, A., Li, S., Tan, L., Sun, Y., Yao, F.: Intelligent measurement and monitoring of carbon emissions for 5Gshared smart logistics. J. Sens. **2022**, 1 (2022)
10. Yao, Q., Li, T., Yan, C., Deng, Z.: Accident responsibility identification model for internet of vehicles based on lightweight blockchain. Comput. Intell. (2022)
11. Luo, Y., Chen, Y., Li, T., Wang, Y., Yang, Y., Yu, X.: An entropy-view secure multiparty computation protocol based on semi-honest model. J. Organ. End User Comput. **34**(10), 1 (2022)
12. Rajasekar, V., Predić, B., Saracevic, M., Elhoseny, M., Karabasevic, D., Stanujkic, D., Jayapaul, P.: Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm. Sci. Rep. **12**(1) (2022)

13. Zhu, S., Liu, Y.: Analysis of human resource allocation model for tourism industry based on improved BP neural network. J. Math. **2022**, 1 (2022)
14. Tong, Z., Ye, F., Mei, J., Liu, B., Li, K.: A novel task offloading algorithm based on an integrated trust mechanism in mobile edge computing. J. Parallel Distrib. Comput. **169**, 185 (2022)
15. Ahmed, I., Zhang, Y., Jeon, G., Lin, W., Khosravi, M.R., Qi, L.: A blockchain- and artificial intelligence-enabled smart IoT framework for sustainable city. Int. J. Intell. Syst. **37**(9), 6493 (2022)

Business, Economics, Finance, Management, Social and Smart Technology

Personalization of Music Streaming Services Using K-means—A Business User-Oriented Model for Spotify Soundtracks Clustering

Ngoc Hong Tran[1], Thanh Duong Cam Nguyen[2], and Dung Hai Dinh[3](✉)

[1] Computer Science, Vietnamese-German University, Thu Dau Mot, Vietnam
[2] Business Administration, Vietnamese-German University, Thu Dau Mot, Vietnam
[3] Business Information Systems, Vietnamese-German University, Thu Dau Mot, Vietnam
dung.dh@vgu.edu.vn

Abstract. In this paper we answer the question: What is the soundtrack to be played next? In order to enhance personalization and assist listeners in discovering music, music streaming services collect a range of data about users' listening habits, such as which songs users play the most frequently, who they share them with, and what time of day they listen to create an unskippable playlist. We build a comprehensive model in MS-Excel, a popular Spread Sheet software of Microsoft, to describe the application and model building process of the K-means Clustering algorithm for recommendation problem of music tracks. We use the dataset containing Spotify music tracks. The results obtained provide a fundamental logic to categorize music samples into several clusters based on their aural characteristics, highlighting the similarities between the songs that have been clustered. Songs that are recommended to a listener are assumed to share some similar musical qualities. This is one of the most important approaches to build a recommendation system, which analyzes preferences and habits of users and recommends the most appropriate products and services. As MS-Excel is familiar with most business users, the template developed in this paper can also be used as a guideline for education and training purposes, as well as a foundation for further applications in various fields and businesses.

Keywords: Unsupervised learning · Clustering problem · K-means clustering · Recommendation system · Spotify tracks

1 Introduction

In recent years, platforms that profit from rising returns have taken control of numerous markets, giving them power and the potential to abuse it. So far, Amazon exploits its own data to disadvantage its suppliers, and Google, whose search algorithm is said to favor properties in which it has a financial stake, see Edelman (2011) or Zhu and Liu (2015). In the music industry, a limited number of services, primarily Spotify and Apple Music, have come to dominate the market as well, and their dominance has prompted questions about potential issues. Together, the two services made over 55%

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 369–378, 2024.
https://doi.org/10.1007/978-3-031-50151-7_35

of the global streaming subscription market in 2019. Around 36% of the market of 304.9 million music subscribers was accounted for by Spotify alone, which was 108.1 million users (Dredge 2019). The trend in the streaming services market continues with expanding market shares of big players, and declining shares of local service providers. For instance, Spotify now has over 40 million songs and is adding roughly 40,000 new ones each week. In this situation, streaming services like Spotify and Apple Music have a significant influence on which songs and artists are successful by the music they choose to promote. On these sites, the "playlists," which are both enlightening lists of music and manipulated tools for listening to them, are the primary means of promotion. Playlists have a wide following, and appearing on important lists has a significant impact on an artist's and song's success (Aguiar and Waldfogel 2018). Given the fact that Spotify can influence success of a song as well as consumption decisions, in this paper we aim to provide a model to implement and understand the clustering process as part of the recommendation system. Our purpose is to let non-technical business users have a foundation to implement extended models and be able to solve business problems with a similar structure. Inspired by the low-code and no-code megatrend in the IT sector, we build an effective and comprehensive model to get interpretable solutions using a standard commercial spreadsheet tool. Taking K = 4 for the K-means clustering algorithm, we separate the dataset into 4 subsets for the purpose of recommending the right product and additional products to customers.

The paper is organized into five sections. Section 1 presents an introduction and problem description. Section 2 reviews the literature and gives background information on recommended playlists and potential causes for bias worries. The dataset of songs is presented in Sect. 3 along with a basic explanation of K-means technique. Section 4 provides details of the clustering process and the result of grouping the songs into 4 playlists. Conclusion and suggestions for further studies are provided in Sect. 5.

2 Review of Relevant Literature

In this digital era, musical items are mostly considered as digital goods that can be accessed and consumed by buyers and users through the internet. However, the ability of consumers to find those new products depends more and more on a small number of platform providers. Prior to digitalization, success hinged on the choices made by program directors at thousands of individually owned radio stations and record stores regarding which songs to highlight on the radio. It has long been recognized that radio promotion and availability in record stores are beneficial. Because of this, record companies have a long history of paying radio stations "payola"—sometimes illegally—for playing their music (Nayman 2012).

Due to the growth of streaming services like Spotify and Apple Music, radio stations and record stores now serve the same functions as streaming platforms in terms of distribution and promotion. Platforms actively choose which songs to include on playlists, as opposed to passively accepting practically all new music. Playlists, which are offered to users as lists of songs, serve two purposes in the promotion of music. First of all, music playlists have the potential to be educational. They serve as tools for music listening, too. When a user clicks on a playlist button, the music in the playlist can be played to them in either random or ranked order. Users of Spotify have the option to "follow" playlists, which makes them visible on the users' home screens. Any user on Spotify is free to make playlists that other users can follow. Even though creating playlists is completely free, Spotify still has control over the popular lists. A number of 866 of the top 1000 Spotify playlists in terms of followers are under the company's control, and their combined followers make up 90% of the followers of these 1000 lists. These calculations are based on the top 1000 playlists listed at Chartmetric's playlist ranking on November 7, 2020.

Despite the fact that numerous supervised automatic music genre classifiers have been put out, they will all rely on the manual classification of the data in the past (Li and Ogihara 2006), (Pye 2000), (McKinney and Breebaart 2003), (Tzanetakis and Cook 2002), (Soltau et al. 1998), (Lee et al. 2009), (Koerich and Poitevin 2005), (Elbir et al. 2018), (Folorunso et al. 2021). Due to this fact, they will not be able to develop along with the data and create new clusters on their own as a result of changing feature values (Shao et al. 2009). An unsupervised technique, on the other hand, would not be dependent on this information and be able to identify the genre of the music samples solely on their auditory qualities. Despite this, few unsupervised methods have been proposed. To obtain a hierarchical structuring music tree, Rauber et al. proposed a growing hierarchical self-organizing map (a popular unsupervised artificial neural network) with psycho-acoustic features such as loudness and rhythm (Rauber et al. 2002). Shao et al. (2004) proposed an unsupervised clustering method that used a hidden Markov model to feed rhythmic content, Mel-frequency cepstral coefficients (MFCCs), linear prediction coefficients, and delta and acceleration values for improvements in feature extraction. By using clustering, commonalities can be uncovered between musical genres without being constrained by their subjectivity, breaking down the barriers created by genre classification. Rewriting music genres is not necessary because these subgenres have unique cultural and historical characteristics that even the most powerful machine learning algorithms cannot fully capture. Instead, grouping based on auditory data can help define more specific subgenres within musical genres and can enhance musical genres.

3 Dataset and Methodology

3.1 Dataset

The data was extracted from the Spotify Web API based on simple REST principles. Its endpoints return JSON metadata about music artists, albums, and tracks, directly from the Spotify Data Catalogue. The dataset has been collected by Hanna Yukhymenko and can be found on Kaggle. There are 20 rows in this data set and 10 different descriptive variables that can be seen in the given list below.

1. Energy - The energy of a song - the higher the value, the more energetic.
2. Danceability - The higher the value, the easier it is to dance to this song.
3. Key - the key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. $0 = C$, $1 = C\sharp/D\flat$, $2 = D$, and so on. If no key was detected, the value is -1 (range: -1; 11)
4. Loudness (dB) - The higher the value, the louder the song.

5. Speechiness - The higher the value the more spoken word the song contains.
6. Acousticness - The higher the value the more acoustic the song is.
7. Instrumentalness - the number of vocals in a song. The closer the value to 1.0, the more instrumental the song is.
8. Liveness - The higher the value, the more likely the song is a live recording.
9. Valence - The higher the value, the more positive mood for the song.
10. Tempo - the overall estimated tempo of a track in beats per minute (BPM).

Figure 1 provides a screenshot of the GUI of our recommendation system. We have a list of 20 records, i.e. songs, and 10 attributes with specific values of each song.

id	artist_name	track_name	danceabili	energy	key	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
1	Olivia Rodrigo	drivers license	0.561	0.431	10	-8.81	0.0578	0.768	1.420E-05	0.106	0.137	143.875
2	Lil Nas X	MONTERO (Call Me By Your Name)	0.593	0.503	8	-6.725	0.22	0.293	0.000E+00	0.405	0.71	178.781
3	The Kid LAROI	STAY (with Justin Bieber)	0.591	0.764	1	-5.484	0.0483	0.0383	0.000E+00	0.103	0.478	169.928
4	Olivia Rodrigo	good 4 u	0.563	0.664	9	-5.044	0.154	0.335	0.000E+00	0.0849	0.688	166.928
5	Dua Lipa	Levitating (feat. DaBaby)	0.702	0.825	6	-3.787	0.0601	0.00883	0.000E+00	0.0674	0.915	102.977
6	Justin Bieber	Peaches (feat. Daniel Caesar & Giveon	0.677	0.696	0	-6.181	0.119	0.321	0.000E+00	0.42	0.464	90.03
7	Doja Cat	Kiss Me More (feat. SZA)	0.764	0.705	8	-3.463	0.0284	0.259	8.920E-05	0.12	0.781	110.97
8	The Weeknd	Blinding Lights	0.514	0.73	1	-5.934	0.0598	0.00146	9.540E-05	0.0897	0.334	171.005
9	Glass Animals	Heat Waves	0.761	0.525	11	-6.9	0.0944	0.44	6.700E-06	0.0921	0.531	80.87
10	MÃ¥neskin	Beggin'	0.714	0.8	11	-4.808	0.0504	0.127	0.000E+00	0.359	0.589	134.002
11	Masked Wolf	Astronaut In The Ocean	0.778	0.695	4	-6.865	0.0913	0.175	0.000E+00	0.15	0.472	149.996
12	Bad Bunny	DÃKITI	0.731	0.573	4	-10.059	0.0544	0.401	5.220E-05	0.113	0.145	109.928
13	Lil Nas X	INDUSTRY BABY (feat. Jack Harlow)	0.741	0.691	10	-7.395	0.0672	0.0221	0.000E+00	0.0476	0.892	150.087
14	Ed Sheeran	Bad Habits	0.807	0.893	11	-3.745	0.0347	0.0451	2.790E-05	0.366	0.537	126.011
15	The Weeknd	Save Your Tears	0.68	0.826	0	-5.487	0.0309	0.0212	1.240E-05	0.543	0.644	118.051
16	BTS	Butter	0.759	0.459	8	-5.187	0.0948	0.00323	0.000E+00	0.0906	0.695	109.997
17	Bruno Mars	Leave The Door Open	0.586	0.616	5	-7.964	0.0324	0.182	0.000E+00	0.0927	0.719	148.088
18	Olivia Rodrigo	deja vu	0.442	0.612	2	-7.222	0.112	0.584	5.700E-06	0.37	0.178	180.917
19	Rauw Alejandro	Todo De Ti	0.78	0.719	3	-3.613	0.0506	0.302	1.960E-04	0.0931	0.336	127.962
20	24kGoldn	Mood (feat. iann dior)	0.701	0.716	7	-3.671	0.0361	0.174	0.000E+00	0.324	0.732	91.007

Fig. 1. Dataset of 20 records and 10 different attributes

3.2 Methodology

The K-means methodology is one of the most widely used methods for forming groups by maximizing qualifying criterion functions, defined either globally (on the entire design) or locally (on the subset of designs) (Patel and Mehta 2011). Given an integer d (songs) and n (song attributes) observations in d dimensions, the task is to choose a cluster of c points which is the number of playlists that would minimize the Mean Squared Distance between each data point and the center to which it belongs. For this problem, no precise polynomial-time solutions are known. However, because solving integer problems with many variables takes a lot of time, clusters are frequently generated using a quick, heuristic method that typically yields good, but not always optimal answers (Jain et al. 1999). One such approach that makes clustering easier is the K-means algorithm. The number of cluster c is initially determined, and their presumed centers are used. Any random item can be used as the starting centroids, or the first k objects in a sequence can also be used. However, if some features have a large size or high variability, these kinds of features will have a significant impact on the clustering outcome. Data normalization would be a crucial preprocessing step in this scenario to scale or manage the variability of the dataset. Then, the three steps below will be performed by the K-means algorithm till convergence. Repetition is done until the stable status is gained, i.e., there is no object left.

- Determine the centroid coordinate.
- Determine the distance of each object to the centroids.
- Group the objects based on minimum distance.

Clustering is often viewed as the most valuable unsupervised learning task because it is assumed that the data sets are unlabeled (Cios et al. 2007). To improve the efficiency of the results, preprocessing is essential before utilizing any data exploration algorithms (Shalabi et al. 2006). One of the preparation steps in data exploration is the normalization of the dataset, where the attribute values are scaled to lie inside a narrow range. For distance metrics, like the Euclidian distance that are sensitive to changes in the magnitude or scales from the attributes, normalization is explicitly required prior to clustering. Normalization stops features with a high number from being prioritized over features with low values. The objective would be to balance the size, scope, and diversity of those elements. The Z-score is a valuable tool when the real lowest and maximum of an attribute X are unknown since it standardizes the values for an attribute X based on its mean and standard deviation. Given a set of raw data Y, the Z-score standardization formula is defined as

$$x_{ij} = Z\left(x_{ij}\right) = \frac{x_{ij} - \overline{x}_j}{\sigma_j} \tag{1}$$

where as

$\overline{x}_j$: sample mean of the jth attribute

σ_j: standard deviation of the jth attribute

The converted variable's mean and variance will both be 0. The original variable's location and scale details are no longer available (Jain and Dubes 1988). The Z-score standardization has certain key limitations, one of which is that it can only be used for global standardization and not within-cluster standardization (Milligan and Cooper 1988). For K-means clustering, given a set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional real vector, K-means clustering aims to partition the n observations into k sets (k $\leq$ n) S = {S_1, S_2,…, S_k} so as to minimize the Within-Cluster Sum of Squares (WCSS):

$$\arg_s \min \sum_{i=1}^{k} \sum_{x_i \in S_i} \left\| x_j - \mu_i \right\|^2 \tag{2}$$

where, μ_i is the mean of points in S_i.

4 Clustering Process and Results

By dividing the data into four playlists, we aim to identify patterns and commonalities among the tracks. The implementation of a Clustering or Classification algorithm can be done with a wide variety of commercial software and data analytics tools, especially the Python programming language for data preparation and Scikit-learn library for model building task. Such library has built-in functions for most state-of-the-art machine learning methods. This paper, for the sake of simplicity and applicability for business users, make use of MS-Excel. We represent the songs and attributes with rows and columns, and utilize functions in the tool without having to write any Macros with VBA programming language. The process of K-means clustering, including metrics calculation and result analysis is described in this section. The first step is standardization.

				Standardization						
id	danceability	energy	key	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
1	-1.087	-1.916	1.046	-1.567	-0.355	2.573	-0.217	-0.621	-1.772	0.348
2	-0.774	-1.344	0.529	-0.438	3.023	0.322	-0.503	1.316	0.694	1.471
3	-0.794	0.730	-1.278	0.235	-0.552	-0.885	-0.503	-0.640	-0.305	1.186
4	-1.067	-0.065	0.788	0.473	1.649	0.521	-0.503	-0.758	0.599	1.089
5	0.291	1.214	0.013	1.154	-0.307	-1.025	-0.503	-0.871	1.576	-0.968
6	0.046	0.189	-1.536	-0.143	0.920	0.455	-0.503	1.413	-0.365	-1.385
7	0.896	0.261	0.529	1.330	-0.967	0.161	1.292	-0.530	0.999	-0.711
8	-1.546	0.460	-1.278	-0.009	-0.313	-1.059	1.417	-0.727	-0.925	1.221
9	0.867	-1.169	1.304	-0.533	0.408	1.018	-0.368	-0.711	-0.077	-1.680
10	0.408	1.016	1.304	0.601	-0.509	-0.465	-0.503	1.018	0.173	0.030
11	1.033	0.182	-0.504	-0.514	0.343	-0.237	-0.503	-0.336	-0.331	0.545
12	0.574	-0.788	-0.504	-2.244	-0.425	0.834	0.548	-0.576	-1.738	-0.745
13	0.671	0.150	1.046	-0.801	-0.159	-0.962	-0.503	-0.999	1.477	0.548
14	1.316	1.754	1.304	1.177	-0.836	-0.853	0.059	1.064	-0.051	-0.227
15	0.076	1.222	-1.536	0.233	-0.915	-0.966	-0.253	2.210	0.409	-0.483
16	0.847	-1.693	0.529	0.396	0.416	-1.051	-0.503	-0.721	0.629	-0.742
17	-0.842	-0.446	-0.245	-1.109	-0.884	-0.204	-0.503	-0.707	0.732	0.483
18	-2.249	-0.478	-1.020	-0.707	0.774	1.701	-0.388	1.089	-1.596	1.540
19	1.052	0.372	-0.762	1.248	-0.505	0.365	3.441	-0.705	-0.916	-0.164
20	0.281	0.348	0.271	1.217	-0.806	-0.242	-0.503	0.791	0.788	-1.353

Fig. 2. Dataset after Z-score standardization

The explanation for the step in Fig. 2 is given in Fig. 3, in which we see the mean and standard deviation of the original values of each of the attributes and thus, the need for standardization.

	danceabilit	energy	key	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
Mean	0,672	0,672	5,950	-5,917	0,075	0,225	0,000	0,202	0,549	133,071
SD	0,102	0,126	3,873	1,846	0,048	0,211	0,000	0,154	0,232	31,079

Fig. 3. Mean and SD of attributes before standardization

For this problem we have to run the iterative process of clustering 4 times in total, i.e. 4 iterations until we get the final clusters. In the first try, we assign the songs to clusters in numerical order, i.e. the first 4 songs are assigned to clusters in a sequence from 1 to 4, then repeat the process, as shown in Fig. 4. This results in a value of Sum of Squares for Error (SSE) of 14,799. More theoretical background about SSE can be found in a lot of literatures, for example the work of Quan (1988).

For the second run, we have an SSE value of 12,984, and although there is a slight improvement, still only 16 songs are correctly clustered. The third run delivers an SSE of 11,422 with 19 correctly clustered songs, and in the final run SSE is 11,074. The final run also gives us a TRUE convergence value since all of the 20 songs are now correctly clustered, thus the algorithm stops. This is summarized in Fig. 5.

The next figure, Fig. 6, provides us a screenshot of the necessary metrics for the iteration process. In C1 to C4 columns, the function SUMXMY2 of MS-Excel is used

Determine the distance of each object to the centroids										
	danceability	energy	key	loudness	speechiness	acousticness	instrumentaln	liveness	valence	tempo
1	-0,020	-0,433	0,633	-0,571	-0,259	0,280	-0,419	-0,782	0,387	-0,254
2	-0,251	0,228	0,116	0,098	0,674	0,232	-0,368	1,180	-0,229	0,286
3	0,453	0,553	-0,710	0,507	-0,519	-0,313	0,695	0,000	-0,029	0,074
4	-0,182	-0,348	-0,039	-0,033	0,104	-0,200	0,091	-0,398	-0,129	-0,106

Fig. 4. Metrics of the first run

1st run										
	danceabilit	energy	key	loudness	speechines	acousticnes	instrument	liveness	valence	tempo
1	-0,02002	-0,43333	0,632642777	-0,57109	-0,25924	0,280144	-0,41864	-0,782	0,38711	-0,25391
2	-0,25052	0,227587	0,116199694	0,098072	0,674449	0,231976	-0,36753	1,18014	-0,22917	0,285651
3	0,452695	0,553278	-0,71010924	0,506504	-0,51911	-0,31253	0,69492	-0,00023	-0,02862	0,074356
4	-0,18215	-0,34754	-0,038733231	-0,03349	0,103906	-0,19959	0,091254	-0,39792	-0,12932	-0,1061
SSE	147,9988									
Converge	FALSE	15								
2nd run										
	danceabilit	energy	key	loudness	speechines	acousticnes	instrument	liveness	valence	tempo
1	-0,098	-0,84521	0,787575702	-1,00242	-0,24737	0,606315	-0,39761	-0,75971	0,089945	-0,07531
2	-0,642	-0,15411	-0,180755079	-0,17165	1,051966	0,503142	-0,47408	1,209296	-0,27371	0,413851
3	0,197	0,795164	-0,342143543	0,82315	-0,65003	-0,56301	0,555926	0,073992	0,196998	-0,18764
4	0,347	-0,59101	0,077466463	-0,47223	0,495478	0,016567	-0,24016	-0,59757	-0,21023	0,036737
SSE	129,846									
Converge	FALSE	16								
3rd run										
	danceabilit	energy	key	loudness	speechines	acousticnes	instrument	liveness	valence	tempo
1	-0,122	-1,07955	0,40024339	-1,36328	-0,31401	1,055139	-0,13502	-0,65378	-0,71375	-0,39835
2	-1,011	-0,42419	-0,30986585	-0,20362	1,591276	0,749527	-0,47408	0,765314	-0,1672	0,67871
3	0,220	0,819657	-0,159236617	0,798467	-0,63433	-0,55208	0,438295	0,178899	0,194308	-0,16346
4	0,851	-0,45398	0,357206466	-0,30621	0,199968	-0,74998	-0,50275	-0,68542	0,591675	0,116569
SSE	114,226									
Converge	FALSE	19,000								
4th run										
	danceabilit	energy	key	loudness	speechines	acousticnes	instrument	liveness	valence	tempo
1	0,118	-1,29072	0,615428008	-1,448	-0,124	1,475	-0,012	-0,636	-1,196	-0,692
2	-1,011	-0,42419	-0,30986585	-0,204	1,591	0,750	-0,474	0,765	-0,167	0,679
3	0,220026	0,819657	-0,159236617	0,798467	-0,63433	-0,55208	0,438295	0,178899	0,194308	-0,16346
4	0,427302	-0,452	0,206577233	-0,50691	-0,0709	-0,61349	-0,50275	-0,69087	0,626821	0,208229
SSE	110,745									
Converge	TRUE	20								

Fig. 5. Metrics of the all four runs

to return the sum of squares of differences of corresponding values in the two arrays, namely the song and the centroid of the cluster, which that song is being assigned to. The figure displays those metrics received in the 1st run, which in turn result in an array for the 2nd run that can be seen on the "Cluster" column. The changes in the order after every run is shown in the four columns on the right-hand side.

As for the final result, Fig. 7 shows that the dataset is clustered into four different groups: songs number 1, 9, and 12 belong to the first cluster, songs number 2, 4, 6, 18 will make up the second cluster, songs number 11, 13, 16, 17 represent the next cluster (cluster 4 in Fig. 7), and all remaining songs belong to the final cluster (cluster 3 in Fig. 7). The further implication for this clustering result is a recommendation mechanism. If, for instance, a listener has listened to song Nr. 6 one time and then repeats it for one or

ID	C1	C2	C3	C4	dist-sq	Cluster	check
1	14,854	21,120	28,532	17,751	14,854	1	TRUE
2	19,678	11,016	26,118	17,076	11,016	2	TRUE
3	10,274	9,412	5,466	6,155	5,466	3	TRUE
4	7,901	7,481	13,761	7,149	7,149	4	TRUE
5	9,814	13,990	7,812	9,152	7,812	3	FALSE
6	13,373	5,870	9,736	9,014	5,870	2	TRUE
7	9,034	13,885	5,231	8,089	5,231	3	TRUE
8	16,156	14,508	8,357	9,218	8,357	3	FALSE
9	5,025	13,159	15,618	8,190	5,025	1	TRUE
10	8,098	4,875	6,819	7,325	4,875	2	TRUE
11	4,770	5,132	4,166	3,082	3,082	4	FALSE
12	10,635	16,411	14,690	10,451	10,451	4	TRUE
13	4,478	12,395	10,477	7,112	4,478	1	TRUE
14	15,520	11,307	8,714	13,398	8,714	3	FALSE
15	19,118	9,879	8,219	13,824	8,219	3	TRUE
16	5,821	12,292	11,224	5,637	5,637	4	TRUE
17	3,036	9,553	8,327	4,158	3,036	1	TRUE
18	21,457	12,046	22,619	17,066	12,046	2	TRUE
19	23,732	23,953	10,287	16,835	10,287	3	TRUE
20	8,454	7,882	6,394	7,354	6,394	3	FALSE
SSE	147,999						
Converge	FALSE	15,000					

1st run	2nd run	3rd run	4th run
1	1	1	1
2	2	2	2
3	3	3	3
4	4	2	2
1	3	3	3
2	2	2	2
3	3	3	3
4	3	3	3
1	1	1	1
2	2	3	3
3	4	4	4
4	4	1	1
1	1	4	4
2	3	3	3
3	3	3	3
4	4	4	4
1	1	1	4
2	2	2	2
3	3	3	3
4	3	3	3

Fig. 6. Changes in the order of songs clustered in 4 groups

id	artist_name	track_name	danceab	energy	key	loudnes	speechi	acoustic	instrumental	liveness	valence	tempo	Cluster
1	Olivia Rodrigo	drivers license	0.561	0.431	10	-8.81	0.0578	0.768	1.420E-05	0.106	0.137	143.875	1
9	Glass Animals	Heat Waves	0.761	0.525	11	-6.9	0.0944	0.44	6.700E-06	0.0921	0.531	80.87	1
12	Bad Bunny	DÃKITI	0.731	0.573	4	-10.059	0.0544	0.401	5.220E-05	0.113	0.145	109.928	1
2	Lil Nas X	MONTERO (Call Me By Your Name)	0.593	0.503	8	-6.725	0.22	0.293	0.000E+00	0.405	0.71	178.781	2
4	Olivia Rodrigo	good 4 u	0.563	0.664	9	-5.044	0.154	0.335	0.000E+00	0.0849	0.688	166.928	2
6	Justin Bieber	Peaches (feat. Daniel Caesar & Giveon	0.677	0.696	0	-6.181	0.119	0.321	0.000E+00	0.42	0.464	90.03	2
18	Olivia Rodrigo	deja vu	0.442	0.612	2	-7.222	0.112	0.584	5.700E-06	0.37	0.178	180.917	2
3	The Kid LAROI	STAY (with Justin Bieber)	0.591	0.764	1	-5.484	0.0483	0.0383	0.000E+00	0.103	0.478	169.928	3
5	Dua Lipa	Levitating (feat. DaBaby)	0.702	0.825	6	-3.787	0.0601	0.00883	0.000E+00	0.0674	0.915	102.977	3
7	Doja Cat	Kiss Me More (feat. SZA)	0.764	0.705	8	-3.463	0.0284	0.259	8.920E-05	0.12	0.781	110.97	3
8	The Weeknd	Blinding Lights	0.514	0.73	1	-5.934	0.0598	0.00146	9.540E-05	0.0897	0.334	171.005	3
10	MÃ¥neskin	Beggin'	0.714	0.8	11	-4.808	0.0504	0.127	0.000E+00	0.359	0.589	134.002	3
14	Ed Sheeran	Bad Habits	0.807	0.893	11	-3.745	0.0347	0.0451	2.790E-05	0.366	0.537	126.011	3
15	The Weeknd	Save Your Tears	0.68	0.826	0	-5.487	0.0309	0.0212	1.240E-05	0.543	0.644	118.051	3
19	Rauw Alejandro	Todo De Ti	0.78	0.719	3	-3.613	0.0506	0.302	1.960E-04	0.0931	0.336	127.962	3
20	24kGoldn	Mood (feat. iann dior)	0.701	0.716	7	-3.671	0.0361	0.174	0.000E+00	0.324	0.732	91.007	3
11	Masked Wolf	Astronaut In The Ocean	0.778	0.695	4	-6.865	0.0913	0.175	0.000E+00	0.15	0.472	149.996	4
13	Lil Nas X	INDUSTRY BABY (feat. Jack Harlow)	0.741	0.691	10	-7.395	0.0672	0.0221	0.000E+00	0.0476	0.892	150.087	4
16	BTS	Butter	0.759	0.459	8	-5.187	0.0948	0.00323	0.000E+00	0.0906	0.695	109.997	4
17	Bruno Mars	Leave The Door Open	0.586	0.616	5	-7.964	0.0324	0.182	0.000E+00	0.0927	0.719	148.088	4

Fig. 7. Result of clustering 20 songs in 4 playlists

some more times, songs Nr. 2, 4, and 18 will then be suggested to that listener because he may also like them.

5 Conclusion and Outlook

The paper presented a clustering model using K-means algorithm which is built in a common spreadsheet software for music recommendation. Recommendation is about helping users enjoy lists of songs that match their preferences. Spotify and other music streaming services gather a variety of information about users' listening habits in order to improve personalization and aid listeners in discovering music, such as which songs users play frequently, with whom they share them, and at what time of day they listen the songs. It is believed that songs which are recommended to a customer share some musical characteristics. The K-means clustering approach was used in this study to group songs into different clusters solely based on their auditory features, highlighting the commonalities between the clustered songs. The dataset was standardized to create consistency between the elements of the song.

It is necessary to mention that this model has a variety of further improvements for use in other business context. Though the example problem seems to be quite easy, an application with thousands of records can require much more computational effort until the convergence is met. Behaviors of similar users, i.e. persons possessing a similar customer profile or taste of music, must also be considered and integrated into the model. One important attribute, which is the artist name, has not been considered. The model can be further developed and considered as a dashboard with appropriate visual interface and pre-configured functional steps. Problems of similar structure but of multi-disciplinary and multi-contextual nature will have higher complexity and solving such cases will become interesting to get valuable insights. Combination with other quantitative methods, such as the Hidden Markov Chain, will also assist the clustering process. A good application would be the clustering problem of students with the available academic records as input attributes. Such a problem setting is promising due to the fact that experimental results on a real dataset can be quite easily collected to demonstrate the effectiveness of the proposed methodology.

References

Rauber, A., Pampalk, E., Merkl, D.: Using psycho-acoustic models and self organizing maps to create hierarchical structuring of music by sound similarity. ISMIR (2002)

Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice Hall, New York (1988)

Lee, C.-H., Shih, J.-L., Yu, K.-M., Lin, H.-S.: Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features. IEEE Trans. Multimedia 670–682 (2009)

Dredge, S.: Musically. Retrieved from https://musically.com/2019/12/09/report-spotify-has-36-market-share-of-music-streaming-subs/ (2019)

Edelman, B.: Bias in search results?: Diagnosis and response. Indian J. Law Technol. **7** (2011)

Elbir, A., Çam, H.B., Iyican, M.E., Öztürk, B., Aydin, N.: Music genre classification and recommendation by using machine learning techniques. In: 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–5. IEEE (2018)

Zhu, F., Liu, Q.: Competing with complementors: an empirical look at amazon.com (2015)

Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. 293–302 (2002)

Milligan, G.W., Cooper, M.C.: A study of standardization of variables in cluster analysis. J. Classif. **5**, 181–204 (1988)

Soltau, H., Schultz, T., Westphal, M., Waibel, A.: Recognition of music types. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1137–1140 (1998)

Jain, A.R., Murthy, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. 265–323 (1999)

Koerich, A.L., Poitevin, C.: Combination of homogeneous classifiers for musical. In: International Conference on Systems, Man and Cybernetics, vol. 1, pp. 554–559 (2005)

Cios, K.J., Swiniarski, R.W., Pedrycz, W., Kurgan, L.A.: Data Mining: A Knowledge Discovery Approach. Springer, New York (2007)

Shalabi, L.A., Shaaban, Z., Kasasbeh, B.: Data mining: a preprocessing engine. J. Comput. Sci. **2**, 735–739 (2006)

Aguiar, L., Waldfogel, J.: Quality predictability and the welfare benefits from new products: evidence from the digitization of recorded music. J. Polit. Econ. **126**(2), 492–524 (2018)

McKinney, M., Breebaart, J.: Features for audio and music classification. (2003). Nayman, L.: Inthesetimes. Retrieved from https://inthesetimes.com/article-/rock-n-roll-payola-dick-clark-and-alan-freed (2012)

Pye, D.: Content-based methods for managing electronic music. In: Proceedings ofthe 2000 IEEE International Conference on Acoustic Speech and Signal Processing, vol. 4, pp. 2437–2440. IEEE (2000)

Folorunso, O., Afolabi, S.A., Owodeyi, A.B.: Dissecting the genre of Nigerian music with machine learning models. J. King Saud Univ. Comp. Inf. Sci. (2021)

Li, T., Ogihara, M.: Toward intelligent music information retrieval. IEEE Trans. Multimedia **8**, 564–574 (2006)

Quan, N.T.: The prediction sum of squares as a general measure for regression di-agnostics. J. Bus. Econ. Stat. **6**(4), 501–504 (1988). https://doi.org/10.1080/07350015.1988.10509698

Patel, V.R., Mehta, R.G.: Impact of outlier removal and normalization approach in modified k-means clustering algorithm. IJCSI Int. J. Comput. Sci. Issues **8**(2), Issue 5 (2011)

Shao, X., Xu, C., Kankanhalli, M.S.: Unsupervised classification of music genre usinghidden Markov model. In: IEEE International Conference on Multimedia, vol. 3, pp. 2023–2026 (2004)

Shao, B., Wang, D., Li, T., Ogihara, M.: Music recommendation based on acoustic features and user access patterns. IEEE Trans. Audio Speech Lang. Process. **17**(8), 1602–1611 (2009)

A Comprehensive Review on Family Budget Management

Rashik Shahriar Akash[1], Mohammad Ullah[1], Radiful Islam[1], Sayed Nahid[1], Ahmed Wasif Reza[2](✉), and Mohammad Shamsul Arefin[1,3](✉)

[1] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
{rashik15-3825,mohammad15-3833,radiful15-3837, nahid15-3849}@diu.edu.bd, sarefin@cuet.ac.bd

[2] Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
wasif@ewubd.edu

[3] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram, Bangladesh

Abstract. In the modern economic era, households have become one of the most significant corporate entities, and personal finances are essential management-related topics. Effective management of personal finances at a family level is really important to address all the financial related issues in an organized way. Any automated budget management system can help a family to utilize their incoming income in an efficient way. In this paper, we try to address the potential researches that were carried out considering the budget management related issues with a strong focus on family budget management.

Keywords: Budget management · Household expenditures · Machine learning · Savings · Income · Expense · Family finance

1 Introduction

Researchers have focused on the issue of analyzing household spending and poverty measurements in recent decades. Income, spending, or consumption data can reflect household wealth indicators. Budget planning has the opportunity to produce more and better insights at a faster rate, adding more value to the sector, thanks to the expanding availability of big data and innovative research methodologies [16]. However, collecting precise income and consumption data for household surveys necessitates a significant investment. Numerous studies have shown substantial regional differences in spending habits and living standards. In reality, expenditure information is a helpful tool for estimating poverty in emerging nations. In general, the GDP is a suitable metric for comparing national wealth.

A fundamental statistical study that is mainly seen used in the majority of countries to accomplish various economic and social objectives is the household budget survey. The tendency of household spending and income, their pattern, and relevant briefing

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 379–391, 2024.
https://doi.org/10.1007/978-3-031-50151-7_36

will be obtained through this study. It is possible to consider creating a prediction model for the consumers' demand because order might significantly influence many different problems.

Budget management is crucial not only at the macroeconomic level, such as in states, but also in families where the needs are often more significant than the income. In addition to managing their spending to compensate for a future income reduction, such as during retirement, households often apply budget management to insure against various types of income risks, including the risk of an unanticipated income decline [19]. Traditionally household budgeting is done by an earning member of the family. Instead of focusing on long-term optimization, short-term management takes over to address the immediate concerns in expenditure patterns. While this approach may be outstanding in making ends meet, long-term calculations and optimizations can improve the family's financial situation by a marginal amount.

Income can be defined as the opportunity for consumption and savings that an entity acquires within a given period and is typically stated in monetary terms. Household income, on the other hand, is defined as the sum of the salaries that are earned by all household members, profits earned due to business or investment, rents and interests, and many other forms of revenues collected over a specific period for all of the members of the household.

Household expenditures habitually rely upon household income, and differences in expenditure patterns highlight the difference in income range between the families. There is also the preferred aspect, as the expenditure is mainly based on preferences. However, expenditure patterns become harder to distinguish from each other when controlling for income.

Expenditure can have a biased effect by the spender that will result in an uneven distribution of income. Other factors like change of lifestyle or change of income may also affect the way money is spent. For this purpose, automation of money management can have a vital role to make life in general more comfortable.

2 Methodology

A systematic review is a survey of the evidence regarding a figured investigation that makes use of purposeful and explicit approaches to identify, choose, and fundamentally evaluate the significant critical analysis. It can be a review of specifics from earlier studies. The following systematic review provides instructions on conducting the survey and adequately analyzing the results.

2.1 Phase 1 - Planning

This section details the procedures used to choose the pertinent publications (such as the search terms and inclusion/exclusion criteria). Prestigious sources for the writing analysis included Springer, ResearchGate, MDPI, Elsevier, IEEE Access, ScienceDirect, and MDPI. The following search phrases were utilized in this audit:

family <OR> financial <AND> income <OR> saving <OR> budget management income <OR> expenditure <OR> expense <OR> budget <OR> saving <OR> financial planning

budget <OR> allocation <AND> machine learning <OR> model

expense <OR> expenditure <OR> budget <OR> income <OR> saving <OR> budget management

(household <OR> (income <OR> expense <OR> expenditure) [AND] (budget <OR> management) [AND] (model <OR> system))

2.2 Phase 2 - Conducting

This section concentrates on the method used to examine the articles.

The papers were rigorously examined for their validity and dependability before being used as the final sample papers for the review. The selected papers were thoroughly inspected to ensure they fulfilled the research's objective.

A few exclusion criteria were used in the literature review on family financial management to limit the selection of final papers for review. First, because the focus of this research is on family budget management, studies that only examined individual budgeting or personal finance were eliminated. In order to make sure the review reflects contemporary trends and practices in family budget management, studies that were conducted more than fifteen years ago were disregarded. Furthermore, due to limited resources for translation, studies that were not published in English were disregarded.

2.3 Phase 3 - Reporting

A total amount of 31 good research papers were picked for review after thorough observation. The research articles are divided into three categories for a systematic evaluation: concept-based, framework-based, and data analysis-based studies. The review is classified to show contributions, work processes, and research errors.

3 Paper Collection

3.1 Identification of Studies

A significant stage was identifying the primary data sources. For the selection of preliminary studies, Google Scholar served as our main search engine. To locate pertinent publications, we also considered several notable academic publishers, including Scopus, IEEE, ScienceDirect, ACM Digital Library, and ResearchGate.

3.2 Study Selection

The process of choosing studies is done in two stages: the first is primary selection, and the second is the final selection.

3.3 Primary Selection

The titles, keywords, and abstracts of the primary sources were first used to choose them; however, when these three criteria proved insufficient, we included the conclusions section to evaluate the process. This phase curated 31 publications, including journals, books, conference papers, and other literature.

3.4 Final Selection

The potential of a research article was assessed based on various criteria, including the scope of the research, methodology, dataset evaluation, key contributions, and future research impact.

3.5 Publication Distribution

A crucial step in preparing a survey report is choosing reliable research articles. Only some research articles published in a particular field are of a high caliber. For our survey to encompass the most recent research and earlier research efforts in seeds categorization, DNN, CNN, and image processing, we chose 5 important research pieces from reputable journals published in five different time frames. Table 1 provides a chronological overview of the projects carried out across various periods for the readers' perusal.

Table 1. Considered papers for review according to time distribution format.

	2008–2016	2017–2018	2019–2020	2021–2022
Conceptual research	Sri et al. [14]			Wasserbacher et al. [16]
	Ha et al. [15]			
Framework based	Adhitama et al. [17]		Rivera et al. [11]	Ismail et al. [13]
	Yadav et al. [21]		Luo et al. [12]	
	Alves et al. [28]		Milewski et al. [10]	
Data analysis	Azadeh et al. [6]	Haque et al. [8]	Antonin et al. [19]	Nigus et al. [4]
	Van Rooij et al. [27]		Lara de Paz et al. [29]	Haque et al. [8]
	Barigozzi et al. [18]	Mohd et al. [30]	Toko et al. [7]	Konşuk Ünlü et al. [20]
	Ahmad et al. [23]		Othman et al. [9]	Zhou et al. [5]
			Jang et al. [1]	Chand et al. [2]

We selected 3 papers on cost prediction, 4 papers on budget allocation and funding and 1 papers on evaluation of models from various publishers, ResearchGate, IEEE, MDPI, ScienceDirect and several others Including ACM.

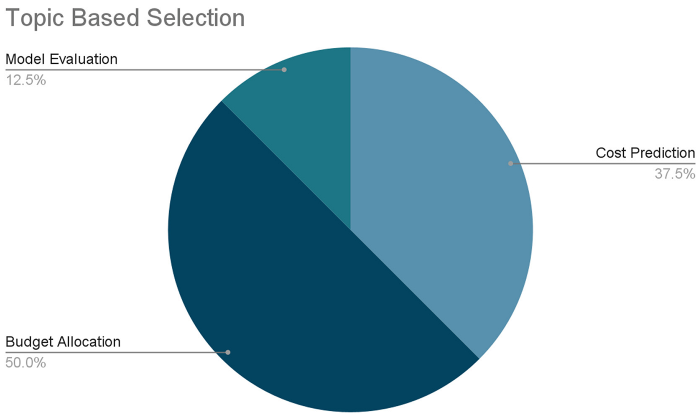


Fig. 1. Topic distribution among selected papers

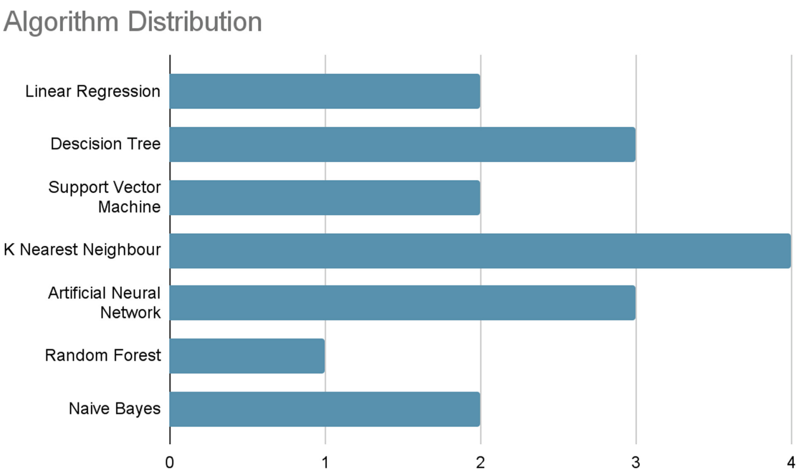


Fig. 2. Algorithm distribution among selected papers

The distribution of topics is shown in Fig. 1.

The distribution of algorithms in the 11 chosen articles, finally, is shown in Fig. 2.

The distribution of publications of the selected documents, according to data sources, is shown in Fig. 3.

4 Analysis Review

In [1], The authors of the study aimed to provide insight into proper management of allocated budgets and their full utilization with the help of machine learning models and various data analysis techniques. They worked on improving the already implemented models that were relevant at that time for the NDIS data. Their analysis aimed to explore the potential of machine learning in the context of NDIS data. There was a demonstration

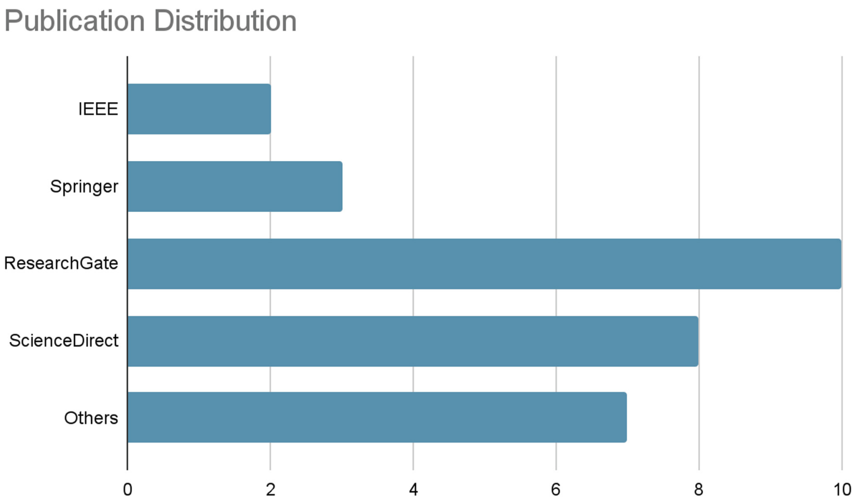


Fig. 3. Publication distribution among selected papers according to data sources

of the overall research framework for this study. Models used for analysis were linear regression, decision trees, support vector machines, and ANNs. The comment was formatted to optimize decision-making tasks using machine learning models or analytic data techniques. There was also a consideration for some improvement in future work that would focus on using the data in a way where the actual expenditure will always be within the limit of the budget allocated (Table 2).

Several recent studies have explored the various combinations of machine learning and analysis to implement in the healthcare management sectors, such as medical information systems, medical technology, and the health insurance sector, with great success. In [2], the study's authors used various machine learning models to improve the budget and resource management for their target sector, the Australian National Disability Insurance Scheme.

The objective was to quantify the budget allocation for each participant to minimize the number of unused funds across NDIS. Specifically, they used NDIS data to examine the ability of Machine Learning Models to predict their fund allocation and the total expenditure of each participant. The analysis includes SVM, ANN, and decision trees. The data show that many NDIS participants left large amounts of unspent funds, but the funds were awarded based on applications reviewed and approved by policymakers. This indicates imbalanced and poor allocation or distribution of funds to the natural and proper needs of single participants and poor choice of the decision by her NDIS regarding the use of funds. Percentage of spending at the individual participant level and b minimizing the number of unused funds across the NDIS.

Using feature selection from the dataset, [4] selected the best features. After that, they divide the data set into parts for training and testing split using the ten-fold cross-validation. Then, using a particular machine learning technique, they applied it to learn about it and tested it. They also evaluated its performance using performance metrics for each machine learning classifier. Naive Bayes (NB), Support Vector Machine

Table 2. Data analysis based

Paper title	Contribution	Dataset	Evaluation
Households income-expenses network simulation [29]	Economic system analyzer with the help of complex network	Mexico's official survey	Interpret a complex system state
Netherlands financial literacy and retirement planning [27]	A suggested strategy for adequately saving and investing for retirement	From questionnaire	Developed a module on financial literacy and plan for it
Household expenditure consumption budget share distributions [18]	Statistical properties of budgeting	Household income and wealth survey obtained from the bank of Italy	Unconditional budgeting as a problem to be solved
Household expenditure prediction [23]	A method for forecasting costs	Household economic integration survey	ANN performed better
Living arrangements of elderly [30]	An analytical study on living agreement of senior citizens	Household income expenditure of Malaysia in 2009	Family tie with the elder members with the effect of income and modernization
Household budget analysis [20]	Huge data was tackled with the help of a regression model	National household budget survey from Turkstat	Demonstrates applicability, monthly consumption expenditure
Links between saving rates, income, and uncertainty [19]	The relationships between French household income and saving rates	Insee household budget survey from 2010 to 11	Extreme income quintiles

(SVM), Gradient Boosting (GB), Extra Tree (ET), Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RF), Logistic Regression (LR) and Ada-Boost (AB) are the algorithms implemented in this case. Except for Gaussian-NB, Linear SVC, and Bernoulli NB, Passive Aggressive Classifier (PAC), the efficiency of the proposed or targeted approach for the accuracy of the training value reaches a higher milestone than 0.80%. Recall, F1-score, and other performance metrics like AUC were also excellent. With a training accuracy of 1.00, the extra tree, decision tree, random forest, and bagging classifiers produce the greatest results. Gradient-Boosting, Ada-Boost, K Neighbors, and Logistic Regression receive scores of 0.9999, 0.9575, 0.9995, and 0.8886, respectively.

In [5], the author used machine learning to more precisely anticipate an OPEX using MLR analysis. Artificial Neural Networks (ANN), k Nearest Neighbors (kNN), Random Forests (RF), and Support Vector Regression are the machine learning approaches used here. The research was divided into two major sections: (1) identifying a relevant component; and (2) assessing the accuracy of the prediction. The primary tool used for the MLR analysis was Minitab. Machine learning analysis was conducted using WEKA, often known as the Waikato Environment for Knowledge Analysis tool. By rescaling the input data from 0 to 1, WEKA automatically sorted and normalized it for all machine learning-based methods. The corresponding prediction MAPE's for the independent test set for the kNN, ANN, RF, and SVR were 11.9%, 8.2%, 9.5%, and 8.5%. Based on the results, kNN had the greatest R2 performance, making it the most accurate method.

In [6], they looked at the relationship between household consumption, income, and LSM. A model for family income spending is also developed. A novel neural network was developed to anticipate and estimate family spending. Four different models have been developed to predict LSM: linear regression, quadratic regression, cubic regression, and GA. In this research, five different forecasting models for household expenditure have been developed. The models were used to assess the utility of the predicted data from the census that was previously incorporated into the model. They used the approach on 18 attributes that are represented by the mean in the dataset. These models are compared using R-square. According to the results, ANN outperformed all other models, with an R2 value of 99.85%. The linear regression model, which had a 99.8% accuracy rate, came in second. The cubic regression's R-square value was 99.6%, while the quadratic's accuracy was 99.6%. The R-square for the GA model is 99.64%.

In [7], coding entails assigning a specific code to each object (or class) type. In the realm of official statistics, this is frequently necessary for the processing of survey data. In this study, they demonstrate how the suggested methodology was applied to the Income and Expenditure of families in Japan. It is simple to process survey items that ask respondents to choose between two or more options while processing survey data. It is challenging to condense general descriptions like profession, industry, and numerous factors affecting family income and expenses. The field of official statistics has seen the development of auto-coding systems. Nearly half of the sectors were covered by the suggested auto-coding method with less than 50% accuracy in most other sectors. In contrast, the rule-based Autocoding System had different coverage for each sector, yet each was reliable and accurate. Both rule-based auto coding and auto coding that uses machine learning offer distinct advantages. A machine learning-based system can classify a sizable amount of the dataset. Although it might be beneficial to design a hybrid system, the precision does not equal the accuracy of a manual coding system.

In [8], they looked into how microfinance affected household income, spending, and savings. They show that when microfinance was used for its intended purpose, ASA clients' income, expenses, and savings considerably rose. As a result, this study concluded that the ASA microcredit program helps rural and urban-disadvantaged houses improve their standard and quality of life through increased income, spending, and savings. Here an empirical model was utilized. They separate it into three halves. In Model 1, the age of the head of the household is found to affect the income of the overall house or family negatively, but this effect is not statistically vital. On the other hand, age is to have a positive, statistically strong, and significant impact, leading to the fact that as the household head becomes older, the effect of it on the household income is impacted more significantly. Household income is positively impacted by the education quality and level of the head and the borrower, a woman, and these effects are statistically significant. In Model 2, The household head's age and the square of that age are not statistically significant and vital. However, it is discovered that the length of ASA participation has a clear relevance and is statistically significant. This shows that household spending rises by 1.13% for every year of participation in ASA. Household heads and the female (borrowers) education levels have a significant and vital correlation. The head of the household's age has a statistically detrimental effect on household savings, according to model 3. Additionally, it is discovered that household savings are positively impacted by the head of the home's age square, and this effect is statistically significant. According to the findings, household savings rise as the head of the household becomes older.

They showed in [9] how data analysis helps figure out what factors contribute to making up the poor category. In this case, less than a quarter of the surveys are used to apply the three models. Since they work with little information, their findings might be off. They invest time and energy into data mining to learn more than is already known. The overall outcomes, driven by the data resource, and the application of those 3 models, provided a clear picture of the analytical potential to manage decision-making to revenue and spending. The results showed that home parameters were the most important determinants of the Family Wellbeing Index. This article did not help determine a lifestyle since it focused on demographic data. More study is needed to build overspending models based on several costs associated with this way of living. The category of costs that generates the most receipts may then be investigated further. As a result, this may inspire communities to be frugal from the get-go rather than waste money because they fear they'll end up in the "overspending" category (Table 3).

There is a model in [12] that users may use to track their finances better. An extensively used supervised learning approach in machine learning is linear regression. The attribute MSE (L2) is utilized as the value. Mean Squared Error (MSE) is the square root of the squared difference between an observation's actual value and its predicted value. The research [10] aims to provide a complete picture of how Philippine income and poverty have changed over time. It also indirectly attempts to clarify the efficacy of anti-poverty policies and initiatives related to income distribution. The model provided in [21] is adaptable, so it may be used with other technologies to improve the model and get us closer to the goal of developing user-specific tools. In this paper, we describe a model that only analyzes user-provided data. When a system that can connect to the internet is installed, it may provide the resident with a wealth of additional financial data, such as the difference in energy consumption between the resident's appliances and those on the market that meet the same criteria. For similar savings and resource preservation, the system may also monitor water and phone use (Table 4).

Using data from 30 semi-structured interviews with males and 122 questionnaires, the authors of [15] set out to better the judgment and budget management procedures. They discovered significant outcomes in managing family finances and making important choices.

Table 3. Framework-based research

Paper title	Contribution	Dataset	Evaluation
Model of performance-based budget planning in public sector entities [11]	To solve the personal financial planning problem and produce better results than those produced by more conventional approaches, this study developed a decision model and used fuzzy goal programming to solve the problem	Private dataset	To demonstrate the efficiency of the strategy, numerical exaASA clients are given. The proposed model is quite effective since it can quickly handle a common problem
Account charting and financial reporting at accounting module on enterprise resource planning using tree traversal algorithm [17]	Adding a sub-module that is a combination of accounts or manually checking the parent and non-parent checkboxes	Private dataset	Unconditional budgeting as a problem to be solved
The impact of education on household income and expenditure inequality [28]	Schooling increases within-level income inequality but not expenditure inequality	The most recent survey on household expenditure from statistics Portugal in 206 is used for the analysis	Using quantile regressions, they have found the importance of education level's impact on within-level driving income disparity

Table 4. Conceptual research

Paper title	Contribution	Evaluation
Budgeting as a methodological basis [14]	Methods for handling one's money and maintaining a household budget	
Machine learning for financial forecasting [16]	An introduction to financial planning and analysis and machine learning	A clear delineation between the responsibilities of forecasting and those of planning

5 Discussion

The rapid increase of prices or inflation has changed and reshaped the lifestyle. Proper family budget management will lead to a more stable and affordable life. The budget section of a family includes both necessary and recreational interests. A balance of both is required to ensure the stability of the family economy. Manual budget management may lead to overspending or underspending in several cases. The primary earner or spender of the family overlooks some instances. Machine learning can significantly help get an unbiased overview of the spending situation. Input for the machine learning algorithm is a detailed dataset that includes comprehensive income to every spending sector. Frequently used prediction algorithms can be used to analyze the dataset. Several relevant works on this have featured algorithms Like ANN, Random Forest, Decision tree, Logistic Regression, and a few others. With the help of those algorithms, an overall prediction of a realistic and doable budget outcome is possible.

This study shows that an effective solution aided by machine learning is yet to be introduced. Most of the existing solutions do not have enough relevancy to be able to help manage the budget of a family. A regression or classification model is simply not enough to be able to perform and generate financial reports that will help maintain a healthy budget distribution for all expenditure sectors. A large amount of training data is needed to train a model that will be able to understand, split, calculate and overview the budget in each spending sector of a family income. There are already such models available at large scale for example openAI and LaMD. A similar but smaller model can be assembled for budget management purposes that may take several inputs and will generate purposeful and intelligent outcomes. Researchers may also examine a wide range of topics in this field in the future. As an example, how social and cultural factors affect family budget management, impact of government policies on family budget management etc.

6 Conclusion

The studies analyzed in this research have shed light on a variety of family budget management issues, such as the value of financial literacy, variables affecting budget management, and techniques for efficient budgeting. The assessment emphasizes that effective budget management depends heavily on financial literacy. According to several studies, families need financial education to improve money management skills, make wise choices, and reach their financial objectives. The studies also emphasize how socioeconomic elements like income, work situation, and educational attainment have an impact on budget management strategies. Budget management is more difficult for low-income families, and they need more assistance to get ahead financially. The analysis also shows that families may manage their finances more effectively by using good budgeting techniques including monitoring costs, establishing financial objectives, and prioritizing spending. The review study emphasizes the value of family budget management in attaining financial success and stability. The research offers insightful information on a range of variables impacting budget management, budgeting techniques, and the significance of financial literacy. The results of this research may be used by policymakers, financial educators, and families to advance financial literacy, boost financial well-being for families, and improve financial management practices.

References

1. Jang, H.: A decision support framework for robust R&D budget allocation using machine learning and optimization. Decis. Support Syst. **121**, 1–12 (2019)
2. Chand, S., Zhang, Y.: Learning from machines to close the gap between funding and expenditure in the Australian national disability insurance scheme. Int. J. Inf. Manag. Data Insights **2**(1), 100077 (2022)
3. Bodendorf, F., Merkl, P., Franke, J.: Intelligent cost estimation by machine learning in supply management: a structured literature review. Comput. Ind. Eng. **160**, 107601 (2021)
4. Nigus, M.: Performance evaluation of classification models for household income, consumption and expenditure data set (2021). arXiv preprint arXiv:2106.11055
5. Zhou, G., Etemadi, A., Mardon, A.: Machine learning-based cost predictive model for better operating expenditure estimations of US light rail transit projects. J. Public Transp. **24**, 100031 (2022)
6. Azadeh, A., Davarzani, S., Arjmand, A., Khakestani, M.: Improved prediction of household expenditure by living standard measures via a unique neural network: the case of Iran. Int. J. Prod. Qual. Manage. **17**(2), 142–182 (2016)
7. Toko, Y., Wada, K., Yui, S., Sato-Ilic, M.: A supervised multiclass classifier as an autocoding system for the family income and expenditure survey. In: Advanced Studies in Classification and Data Science, pp. 513–524. Springer, Singapore (2020)
8. Haque, A.C., Das, A., Rahman, A.: The effectiveness of micro-credit programmes focusing on household income, expenditure and savings: evidence from Bangladesh. J. Compet. **9**(2) (2017)
9. Othman, Z.A., Bakar, A.A., Sani, N.S., Sallim, J.: Household overspending model amongst B40, M40 and T20 using classification algorithm. Int. J. Adv. Comput. Sci. Appl. **11**(7) (2020)
10. Milewski, R., Tomasiewicz, M.: Model of performance-based budget planning in public sector entities. Sci. J. Milit. Univ. Land Forces **51** (2019)
11. Rivera, J.P.R.: Estimating Gini coefficient and FGT indices in the Philippines using the family income and expenditure survey. J. Poverty **24**(7), 568–590 (2020)
12. Luo, Y.: Resident consumption expenditure forecast based on embedded system and machine learning. Microprocess. Microsyst. **83**, 103983 (2021)
13. Ismail, R., Abu Bakar, N.: The relationship between income, expenditure and household savings in Peninsular Malaysia. Malays. J. Consum. Fam. Econ. **15**, 168–189 (2012)
14. Sri, Y.B., Sravani, Y., Surendra, Y.B.S., Rishitha, S., Sobhana, M.: Family expenditure and income analysis using machine learning algorithms. In: 2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Dec 2021, pp. 1–5. IEEE
15. Ha, N.N.: Household budget management and decision-making in the family in the Red River Delta, Vietnam. Vietnam J. Fam. Gend. Stud. **14**(2), 17–29 (2019)
16. Wasserbacher, H., Spindler, M.: Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. Digit. Finance 1–26 (2021)
17. Adhitama, M.A., Sarno, R.: Account charting and financial reporting at accounting module on enterprise resource planning using tree traversal algorithm. In: 2016 International Conference on Information & Communication Technology and Systems (ICTS), Oct 2016, pp. 20–25. IEEE
18. Barigozzi, M., Alessi, L., Capasso, M., Fagiolo, G.: The distribution of household consumption-expenditure budget shares. Struct. Change Econ. Dyn. **23**(1), 69–91 (2012)
19. Antonin, C.: The links between saving rates, income and uncertainty: an analysis based on the 2011 household budget survey. Econ. Stat. **513**(1), 47–68 (2019)

20. Konşuk Ünlü, H.: A new composite lognormal-Pareto type II regression model to analyze household budget data via particle swarm optimization. Soft Comput. **26**(5), 2391–2408 (2022)
21. Yadav, S., Malhotra, R., Tripathi, J.: Smart expense management model for smart homes. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), Mar 2016, pp. 544–551. IEEE
22. Rashid, N.K.A., Nasir, A., Mustapha, N.H.N., Kamil, N.F.: Analysis of income and expenditure of households in the east coast of Peninsular Malaysia. J. Glob. Bus. Econ. **2**(1), 59–72 (2011)
23. Ahmad, Z., Fatima, A.: Prediction of household expenditure on the basis of household characteristics. Islam. Ctries. Soc. Stat. Sci. **351**
24. Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., Alyaman, M.: Poverty classification using machine learning: the case of Jordan. Sustainability **13**(3), 1412 (2021)
25. ChiangLin, C.Y., Lin, C.C.: Personal financial planning based on fuzzy multiple objective programming. Expert Syst. Appl. **35**(1–2), 373–378 (2008)
26. Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., Evanoff, D.D.: Financial literacy and financial planning: evidence from India. J. Hous. Econ. **27**, 4–21 (2015)
27. Van Rooij, M.C., Lusardi, A., Alessie, R.J.: Financial literacy and retirement planning in the Netherlands. J. Econ. Psychol. **32**(4), 593–608 (2011)
28. Alves, N.: The impact of education on household income and expenditure inequality. Appl. Econ. Lett. **19**(10), 915–919 (2012)
29. Lara de Paz, J., Flores de la Mota, I., Policroniades Chipuli, G., Shirai Reyna, S.: Households income-expenses network simulation. In: European Modeling & Simulation Symposium, Sept 2019, pp. 210–217. CAL-TEK Srl
30. Mohd, S., Senadjki, A., Mansor, N.: Living arrangements of elderly: evidence from household income expenditure survey. J. Popul. Ageing **10**(4), 323–342 (2017)

Analysis of Tweets for Tourism Trends: A Case Study of Kathmandu, Nepal

Deepanjal Shrestha[1,2](✉), Sujan Tamrakar[3], Ghanshyam Subedi[3], Ravindra Boojhawon[4], and Purna Bahadur Khand[1]

[1] School of Business, Pokhara University, Pokhara, Nepal
deepanjal@hotmail.com, purnabahadurkhand@gmail.com
[2] Nanjing University of Aeronautics and Astronautics, Nanjing, China
[3] Gandaki College of Engineering and Science, Pokhara University, Pokhara, Nepal
tamrakarsujaan061@gmail.com, subedi.ghanshya_m@gmail.com
[4] University of Mauritius, Moka, Mauritius
r.boojhawon@uom.ac.mu

Abstract. Social media has gained importance in social and business sectors. It is capable of creating value with its rich information content and real-time dissemination of information from live human sensors. The role of these social sites cannot be ignored in information-intensive systems like the tourism business systems. This study presents a case study of Twitter trend analysis of Kathmandu City as a tourism destination in Nepal. The work collected around 30,000 tweets using Twitter's application programming interface (API), and analyzed them using data mining tools and techniques. The analysis revealed that Twitter users in Kathmandu were highly engaged in discussing tourism-related activities and events. The work identified several distinct trends in the Twitter data, including topics such as the earthquake, Nepali food, heritage sites, and tourism Points of Interest (POI). The work also observed trends in the use of hashtags, sentiments related to Twitter content, time series analysis, and information clusters. This study is important to understand the position of Kathmandu in online sites as a tourism destination and the opinions tourist carry for it. This can help the tourism businesses, governing bodies and associated stakeholders to utilize the knowledge to improve the overall image of the city as a tourism destination. It also contributes to the existing literature on social media analysis and provides a foundation for further research in this area.

Keywords: Tweets · Hashtags · Tourism · Information clusters · Sentiment analysis · Kathmandu · Nepal

1 Introduction

Online social media has gained a lot of popularity and usage in modern times [1]. These platforms are powerful means to create and disseminate information in real-time and in varied formats including audio, video, text, photos, etc. [1]. There are various social sites available today which include Facebook, Twitter, WeChat, YouTube, etc. and are capable

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 392–402, 2024.
https://doi.org/10.1007/978-3-031-50151-7_37

of connecting millions and millions of users on daily basis [2]. These sites help people to connect to one another, share ideas and information, market product and services, and provide participation and open dialogue in politics and social scenarios. The role of social sites cannot be ignored today by any business house or governing body if they want to reach a larger audience and create a good impact.

Twitter is one of the most powerful social networking as well as a micro-blogging site. It is a powerful platform that has information on the latest trends, user opinions, and sentiments for various topics of interest. Twitter data is rich and has different attributes that can help any business to establish and grow [3]. The data can provide valuable insights into people and their preferences in terms of products and services. Twitter data is very important for industries like tourism, which are information intensive and rely on tourist opinions. With millions of daily active users sharing their thoughts and experiences in real time, Twitter provides a wealth of data that can help tourism businesses and marketers gain valuable insights into consumer behavior and preferences [3]. By analyzing Twitter trends, tourism professionals can identify popular destinations, activities, and experiences and understand how travelers feel about them. This information can help businesses tailor their marketing strategies, improve customer experiences, and make informed product development and innovation decisions. In this way, Twitter trend analysis can provide a competitive edge for tourism businesses looking to stay ahead of the curve in an ever-changing industry. Further, the data on tourism can also be utilized by countries dependent on tourism to develop their economy. The data gathered through this site is helpful for tourism development and governance.

2 Literature Review

Different scholars have carried out research on Twitter at different times with subjects ranging from trend analysis to real-time event detection, geospatial analysis, recommendations and the use of tweets for business and social purposes. Some notable work is mentioned in Table 1.

The studies of social and online media in Nepal are limited to a few authors and only very few have worked on the utilization of Tweets in Nepal. The study of area-specific Tweets and their trend is missing at large. These studies suggest that social media can provide valuable insights into various aspects of Nepal, such as tourism and governance. Sentiment analysis, prediction of tweet popularity, and real-time event detection are effective tools for analyzing social media data. Data analytics and content analysis are also important for gaining insights into tourism-related data and social media usage This work adds up as a vital knowledge base for Twitter-based studies in Nepal.

3 Research Framework

The work uses Python, Twitter scrapper module and Twitter API to download data from Twitter based on the keyword "Kathmandu" for a period of 12 years. The data was obtained as a CSV file and is further processed to make it rich and error-free. The conceptual framework of the research is shown in Fig. 1.

Table 1. Literature review on Twitter and its domain of study

Author and year	Area	References
Peng et al. (2015)	Trending sentiment-topic detection on Twitter	[4]
Hasan et al. (2018)	Real-time event detection from the Twitter data stream using the Twitter	[5]
Mahdikhani (2022)	Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic	[6]
Yakobi et al. (2022)	Review of the social media analytics framework for citizen relationship management	[7]
Wenan et al. (2022)	Analysis and evaluation of TripAdvisor data: a case of Pokhara, Nepal	[8]
Devkota and Miyazaki (2019)	Using volunteered geographic information and nighttime light remote sensing data to identify tourism areas of interest	[9]
Devkota and Miyazaki (2018)	An exploratory study on the generation and distribution of geotagged tweets in Nepal	[10]
Bhatt and Pickering (2022)	Destination image of Chitwan National Park, Nepal: insights from a content analysis of online photographs	[11]
Shrestha and Jeong (2015)	ICT, e-governance challenges and factors affecting practices in Nepal	[12]

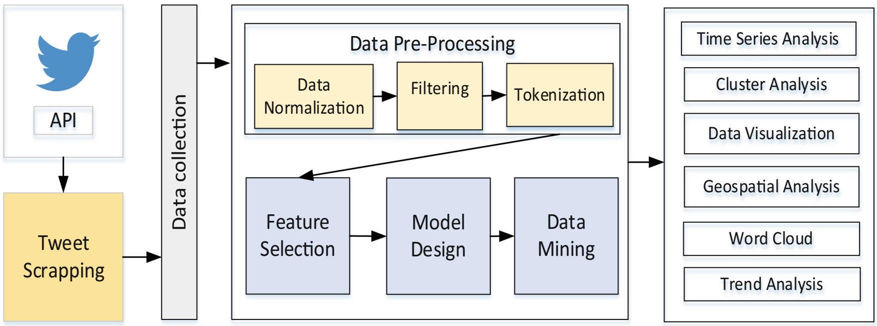


Fig. 1. Conceptual framework of the research

3.1 Data Collection

The data was collected for a period from 2008 to 2020 using the Twitter scrapper module for Python 3.3 version, Twitter API, and supporting libraries. The module returned around 30000 Tweets for the keyword Kathmandu that was further preprocessed to avoid missing values, noise, outliers, errors, and inconsistency. An algorithm to download the Tweets is shown below:

Algorithm 1. Downloading tweets	
Define *twitter scraper module and import function*	
import *data frame library*	
import *date time library*	
#Define	
	twitter_credentials.CONSUMER_KEY
	twitter_credentials.ACCESS_TOKEN
	begin_date = dt.date(yyyy/m/d)
	end_date = dt.date(yyyy/m/d)
	#limit = n
	#lang = 'english'
tweets = query_tweets("keyword", begindate = begin_date, enddate = end_date)	
df = pd.DataFrame(t.__dict__ for t in tweets)	
#Options	
	#limit = limit,lang = lang

3.2 Data Preprocessing

Data cleaning is done by removing irrelevant information such as URLs, hashtags, and user mentions. This was done using specialized libraries like NLTK. The tokenization is carried out to identify individual words and hashtags to analyze the text at a more granular level. Stop word removal and stemming and lemmatization is carried out.

4 Data Analysis and Findings

The Twitter data set was further analyzed using time series, clustering, geospatial and word processing techniques. Feature selection and data segregation was carried to perform better visualization models of the data. Finally, sentimental analysis was carried to know the sentiments of people for Kathmandu valley.

4.1 Inferential Statistical Analysis

4.1.1 Correlation Analysis

The linear relationship between sentiment attributes was measured using the Pearson correlation. It was noted that the compound value of positive sentiments is 0.9 indicating a strong correlation between neutral and positive attributes. The other values of correlation are negative ranging from − 0.01 to − 0.826 showing weak to strong negative correlation of sentiments for Kathmandu city. As the overall compound value was higher, Kathmandu city has strong positive sentiments as shown in Table 2.

Table 2. Pearson correlation

1	0.9	Compound	Pos
2	– 0.826	Neu	Pos
3	– 0.621	Compound	Neu
4	– 0.374	Neg	Neu
5	– 0.33	Compound	Neg
6	– 0.01	Neg	Pos

4.1.2 Term Frequency-Inverse Document Frequency (TF-IDF) Analysis

TF-IDF (Term Frequency-Inverse Document Frequency) This analysis was carried to study the importance of a term within a document or a corpus. It is a numerical statistic that reflects the importance of a term within a document or a corpus and is computed as below:

$$TF\text{-}IDF\ is\colon TF\text{-}IDF(t, d, D) = TF(t, d) * IDF(t, D) \tag{1}$$

where:

- t is a term (word or phrase), d is a document
- D is the corpus (collection of documents)
- *TF(t, d)* is the term frequency of t in d
- *IDF(t, D)* is the inverse document frequency of t in D, calculated as
- *IDF(t, D) = log(N/n_t)*

where:

- N is the total number of documents in the corpus
- *n_t* is the number of documents in the corpus that contain the term t.

Table 3 represents the top 10 words with TF-IDF scores and it was seen that Kathmandu has the highest score of 0.571144 followed by Nepal with 0.072862 and flight as the third with 0.034713. It was interesting to note that words like an earthquake (0.0143283) and Nepal_earthquake (0.00217) had a prominent presence. The table provides an understanding of the top informative words in context or in relation to Kathmandu Tweets.

4.2 Time Series Analysis

A time series analysis of the data collected from 2008 to 2020 was performed for the keyword Kathmandu as shown in Fig. 2. The time series analysis shows that the number of tweets has grown in the context of Kathmandu with certain gaps in between. The gaps are seen after the occurrence of the earthquake [13] and during the Covid-19 period. It can also be inferred that Nepalese are less accustomed to Tweets due to less posting of tweets during this period of time.

Table 3. TF-IDF top 10 values from tweet corpus

SN	Word	TF-IDF
1	Kathmandu	0.571144
2	Nepal	0.072862
3	Flight	0.034713
4	Pokhara	0.020792
5	Weather	0.014826
6	Earthquake	0.014283
7	Travel	0.006147
8	Deals	0.003616
9	Thamel	0.002712
10	News	0.002531

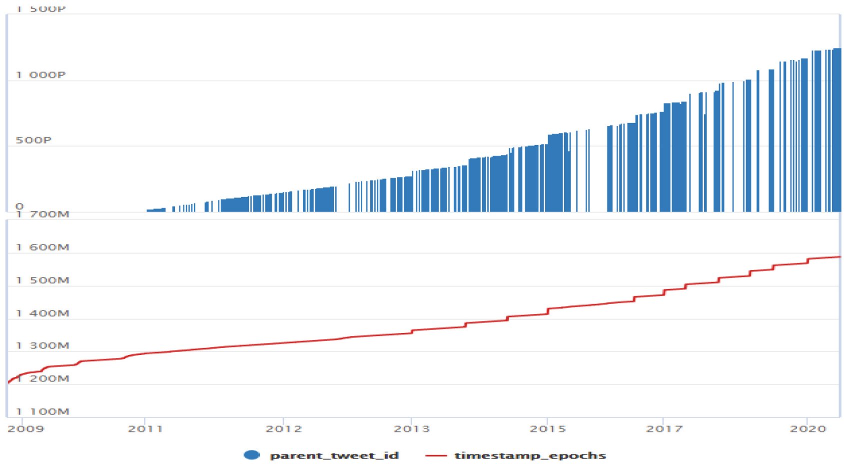


Fig. 2. Time series plot of parent tweet and timestamp epochs.

The data shown in Fig. 3 provides an insight into the behavior of Twitter users and shows that during the early period of 2008–2013 the number of tweets was limited from 1 M to 100 M. The tweets are seen increasing 1 G to 100 P in during 2016 to end of 2018 and are again seen growing from January 2020.

4.3 Information Cluster and Sentiment Analysis

A cluster analysis was for the collected data set to understand the emotions attached to the city. It was found that surprise was the most used word with 17205 self-combined occurrences followed by a combination of joy and surprise with 121 combined occurrences. The other words that were seen in combination with the earlier words were

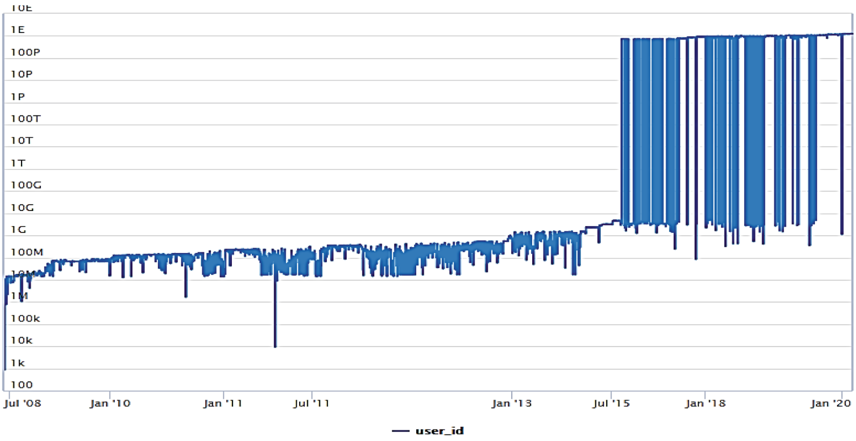


Fig. 3. Representing data of user_id vs time

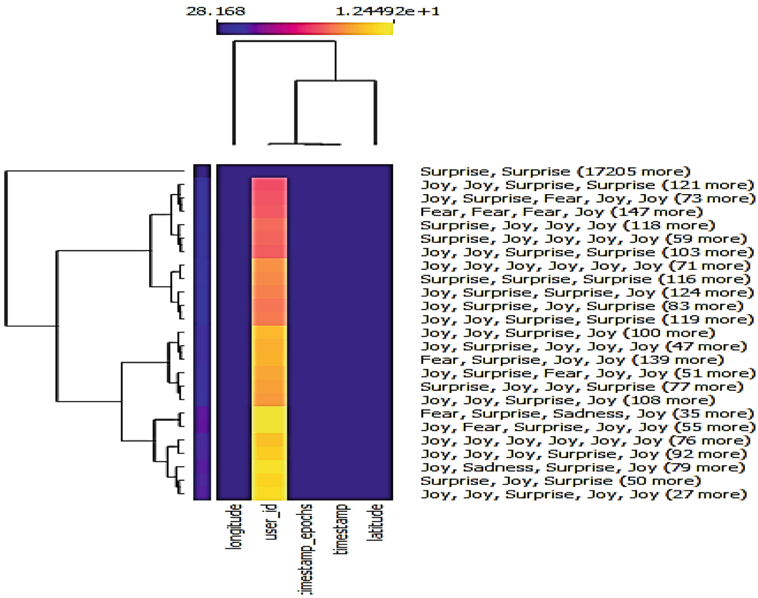


Fig. 4. A heat map depicting a cluster of emotions for tweet attributes.

fear. A heat map for the combination with user_id, time stamp, timestamp_epochs, and geotagged data (latitude and longitude) is shown in Fig. 4. A detailed analysis of the emotions embedded in the tweets over time is shown in Figs. 5 and 6. It was seen that users have expressed different emotions during different time periods. The period from 2011 to 2018 has surprised as the dominant emotion, while 2015 accounts for anger, sadness, fear, and disgust. The emotion for the time period 2015 can be understood with

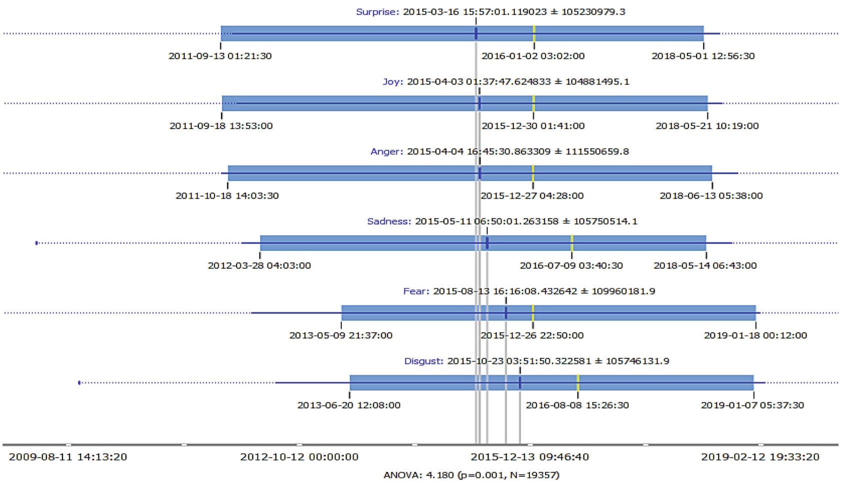


Fig. 5. A boxplot depicting user emotions over time.

negative emotions as Nepal was hit by an earthquake in the same year [13] and many tourism services were badly hit.

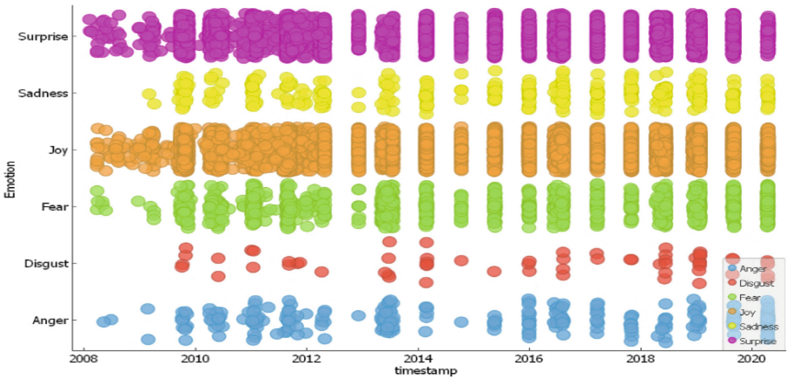


Fig. 6. Representing the sentiments in tweets over time for Nepal tourism.

Similarly, a word cloud was obtained using the corpus of hashtag tweets and it was seen that the most frequent words are Kathmandu, Nepal, Pokhara, Lumbini, travel, paragliding, Annapurna, visitNepal, 2020, Buddha. The word cloud trend shows that more than 85% of the words in the visualization talk about tourism trends like visit Nepal or a tourism point of interest like Lumbini, Pokhara, or Kathmandu, or about food and culture like momos and coffee. The word cloud also show trends like visit Nepal, travel Nepal and natural disasters like earthquake and pandemic like Covid 19. Sporting events like mountain bike, paragliding, trekking and hiking are also present with other

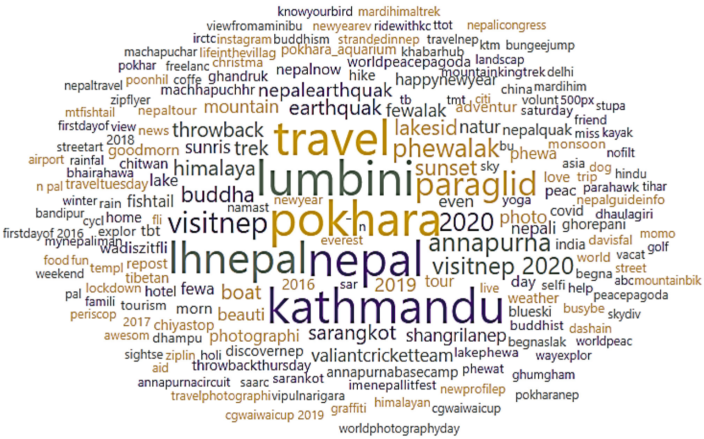


Fig. 7. Representing the word cloud extracted from the tweet corpus.

similar types of tourism products, services and information. Figure 7 represents all the prominent words of the word cloud obtained from the hashtag corpus.

4.4 Geo Spatial Analysis

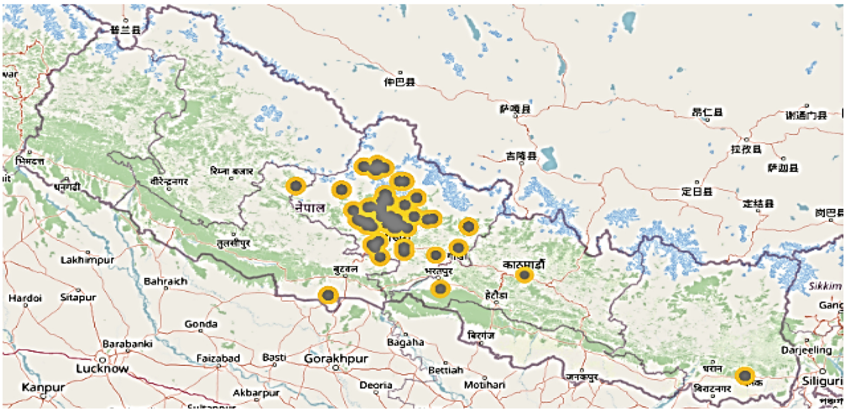


Fig. 8. Geotagged tweets obtained for search keyword Kathmandu

The geotagged tweets collected for the keyword "Kathmandu" are displayed in Fig. 8, which reveals a high concentration of tweets in and around the city. However, it is noteworthy that some tweets originated from other popular locations like Pokhara, Lumbini, Namche Bazar, Mustang, and alike, indicating other popular tourist destinations in Nepal. In addition to Kathmandu, we have also plotted geolocation data for Lumbini, Birendranagar, Janakpur, and Pokhara in the same figure. To provide a more comprehensive

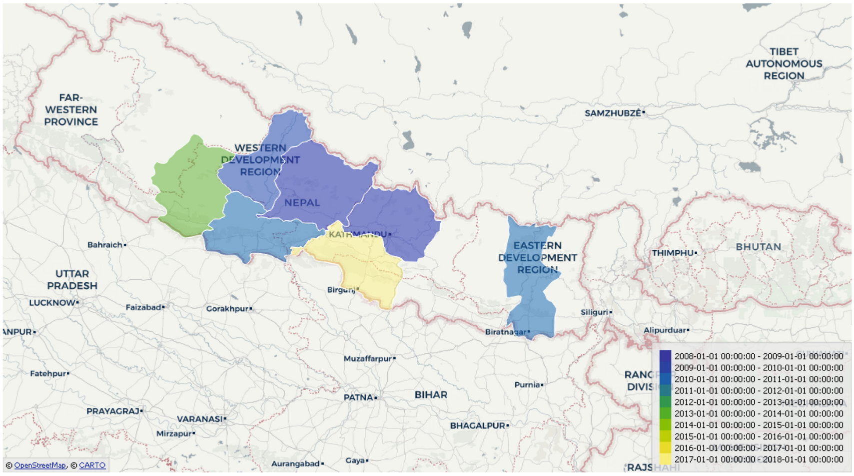


Fig. 9. Regions of origin of tweets with timestamp information.

picture, we have included a time-series plot in Fig. 9 that indicates the regions from where the tweets originated, along with the time of origin. Our analysis shows that tourists tend to flock to popular destinations in Nepal, with the majority of tweets generated from September to late February.

5 Conclusion

In conclusion, the paper "Analysis of Tweets for Tourism Trends: A Case Study of Kathmandu, Nepal" provides a comprehensive analysis of Twitter data to identify tourism trends in Kathmandu. The study uses data mining techniques to extract relevant information from tweets and analyze it to draw meaningful insights. It was seen that Twitter users have grown over the year for Kathmandu Nepal and have used it as a primary source to talk about tourism products, services, places, trends, food, and disaster situations like earthquakes and Covid-19. The overall sentiments of the users for the city are neutral to positive with more emotional content being posted in later years. The geotagged tweets show that users have mentioned Kathmandu while traveling in other locations, showing the positive correlation of Kathmandu with other tourism destinations. The study finds Kathmandu as a central location for connecting other tourism destinations and is known for cultural tourism, and religious tourism, and identifies specific attractions and events that are driving tourism to the city. The overall, study provides a useful example of how social media data can be used to gain insights into tourism trends and highlights the potential value of this approach for future research in the field.

References

1. Shrestha, D., Rajkarnikar, N., Shrestha, D., Dhakal, R., Jeong, S.R.: Impact of social network sites in the life of university students: a case study of Pokhara University, Nepal. Int. J. Innov. Res. Comput. Sci. Technol. (IJIRCST) **9**(4) (2021)

2. Wenan, T., Shrestha, D., Shrestha, D., Rajkarnikar, N., Jeong, S.R.: The role of emerging technologies in digital tourism business ecosystem model for Nepal. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_105
3. Devkota, B., Miyazaki, H., Pahari, N.: Utilizing user generated contents to describe tourism areas of interest. In: 2019 First International Conference on Smart Technology & Urban Development (STUD), Chiang Mai, Thailand, pp. 1–6 (2019). https://doi.org/10.1109/STUD49732.2019.9018810
4. Peng, B., Li, J., Chen, J., Han, X., Xu, R., Wong, K.F.: Trending sentiment-topic detection on Twitter. In: Gelbukh, A. (ed.) Computational Linguistics and Intelligent Text Processing. CICLing 2015. Lecture Notes in Computer Science, vol. 9042. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18117-2_5
5. Hasan, M., Orgun, M.A., Schwitter, R.: Real-time event detection from the Twitter data stream using the TwitterNews+ framework. Inf. Process. Manag. **56**(3), 1146–1165 (2019). https://doi.org/10.1016/j.ipm.2018.03.001
6. Mahdikhani, M.: Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic. Int. J. Inf. Manag. Data Insights **2**(1), 100053 (2022). https://doi.org/10.1016/j.jjimei.2021.100053
7. Yakobi, K., Scholtz, B., vom Berg, B.W.: An expert review of the social media analytics framework for citizen relationship management. In: Marx Gómez, J., Lorini, M.R. (eds.) Digital Transformation for Sustainability. Progress in IS. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-15420-1_17
8. Wenan, T., Shrestha, D., Gaudel, B., Rajkarnikar, N., Jeong, S.R.: Analysis and evaluation of TripAdvisor data: a case of Pokhara, Nepal. In: Vasant, P., Zelinka, I., Weber, G.W. (eds.) Intelligent Computing & Optimization. ICO 2021. Lecture Notes in Networks and Systems, vol. 371. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93247-3_71
9. Devkota, B., Miyazaki, H., Witayangkurn, A., Kim, S.M.: Using volunteered geographic information and nighttime light remote sensing data to identify tourism areas of interest. Sustainability **11**, 4718 (2019). https://doi.org/10.3390/su11174718
10. Devkota, B., Miyazaki, H.: An exploratory study on the generation and distribution of geotagged tweets in Nepal. In: 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Kathmandu, Nepal, pp. 70–76 (2018). https://doi.org/10.1109/CCCS.2018.8586827
11. Bhatt, P., Pickering, C.M.: Destination image of Chitwan National Park, Nepal: insights from a content analysis of online photographs. J. Outdoor Recreat. Tour. **37**, 100488 (2022). https://doi.org/10.1016/j.jort.2022.100488
12. Shrestha, D., Jeong, S.R.: Challenges and factors affecting E-governance practices in Nepal. In: 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Kathmandu, Nepal (2015)
13. Rafferty, J.P.: Nepal earthquake of 2015. Encyclopedia Britannica, 3 Jan 2023. https://www.britannica.com/topic/Nepal-earthquake-of-2015. Accessed 10 Mar 2023

Ergonomics Design: A Smart Concept of Indian Kitchen

Amit Bhatia[1], Sandeep Singla[1](✉), and Aarti Bansal[2]

[1] Department of Civil Engineering, RIMT University, Mandi Gobindgarh, Punjab, India
dr.sandeepsinglaz@gmail.com

[2] Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
aarti.bansal@chitkara.edu.in

Abstract. The kitchen is one of the important areas of where food is cooked and utensils are washed. The ergonomics of kitchen plays an important role in reducing muscle pains of women. Using ergonomic design, the kitchen area can be made fatigue free and reduce unnecessary movements where as an unplanned kitchen construction reduces work efficiency and thereby increasing more time and effort. The current research reviews existing work conducted in improving design of kitchen in order to reduce the muscle pain and other disorders. The studies are based on conventional and modular kitchens designs.

Keywords: Ergonomics · Kitchen · Postures · Smart infrastructure

1 Introduction

Unlike the other rooms of the home, the kitchen is a workplace of the housewife. The tasks of cooking, serving and cleaning up the dishes after lunch and dinner are done in the kitchen. A housewife spends most of the time in the kitchen. Therefore, to make the kitchen a better place for the housewife/cook is a thoughtful subject. Hence, the goal of Ergonomics is to make this work more comfortable, faster, more delightful and less tiring for the housewives, and this can become possible when the interface between them and the other objects which they need to accomplish their task gets easier. The movement capabilities of the body are like the constant parts of the equation. The actions of human beings can't be altered according to the environment of the kitchen. However, the kitchen environment can be altered to make it more efficient and comfortable. The alterations of kitchen environment should be based upon movement abilities of human beings. The tasks of the kitchen will become easier if the physical movements i.e. walking, twisting the body, lifting the hands, bending, stretching and leaning are reduced.

The good designers of the kitchen design the interior of the kitchen by applying the policy of ergonomics which makes the kitchen a flexible place where everyone can do the work without wastage of time and energy and high efficiency. If the work is less time consuming and less energy consuming, then it will cause low fatigue. Hence, the interior of the kitchen should be designed in this way, that it will not take much time

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 403–409, 2024.
https://doi.org/10.1007/978-3-031-50151-7_38

and energy in the movement activities of bending, twisting, walking, lifting, etc. so that the housewives will get more time for cooking and enjoying the lunch and dinner with family.

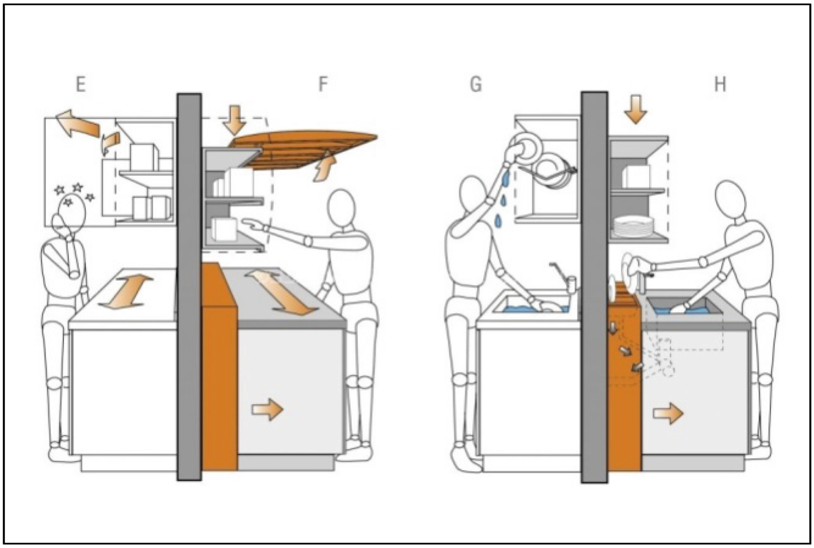


Fig. 1. Kitchen ergonomic design

By making certain modifications like appropriate placement of microwave, dishwasher and kitchen sink the kitchen ergonomics can be improved significantly as shown in Fig. 1.

According to the early studies of 1920, the basic kitchen triangle in a kitchen can reduce the household chores by 70% from cooking the food to washing up the dishes. Usually, from preparing the food to cooking it, everything is done at a work surface - countertop or table - that is sufficient for doing the task in the kitchen. Hence, the ingredients and the utensils i.e. Oven, microwave, pots, pans, etc. should be placed just on the right-hand side for easy availability. After cooking the meal, the other utensils in which the meal is served should be available easily. Ultimately, after taking the meal, the cleaning up of dishes in the sink is done. Again the dishwasher and other things needed for cleaning up the dishes should be available easily.

Figure 2 shows knee space under sink which is designed as per principles of ergonomics. The area of knee space in the kitchen helps the person sitting on a wheelchair to do the tasks. The housewives can do the tasks of cooking, preparing and cleaning up while sitting on a stool comfortably. The serving cart can also be put under this area and the setting or cleaning off a table can be done in only one time rather than many.

Fig. 2. A knee space under a sink

2 Related Work

Statistiska et al. [1] In the kitchen, many kinds of physical work are done, and in this case, the housewives/cooks can have some muscular disorders in their bodies. This study aims at reducing the chances of muscular disorders. The goal of this study is to make them aware and efficient to use the ergonomics in the kitchen. This study shows the "intervention process and evaluates its feasibility with regard to the elements of the intervention process and available resources and support" [1].

Hirvonen et al. [2] This study shows the impacts of the intervention on the housewives'/cooks' knowledge and awareness about ergonomics. This also shows their expectations from the ergonomics and how it has helped them by reducing the household chores as well as the chances of muscular disorders.

Haukka et al. [3] In the study, it was found out that the workers were chiefly women (96%) around 46 years of age (range 19–63). The average timing in the domestic work of the kitchen was 17 years (range 0–40). 87% of them suffered from muscular pain. 71% of them suffered from neck pain; 50% of them suffered from lower back pain, and 49% of them suffered from the hands' pain. "The 59 intervention kitchen belonged to schools (n ¼ 43), kindergartens (n ¼ 10), nursing homes and geriatric service centers (n ¼ 5), and one was a central kitchen. The total number of workers was 263. Eighty-six percent (n ¼ 227) of the employees in the same kitchen throughout the intervention phase" [3].

Mahajan et al. [4] The study of ergonomics is very important. This is applied in sports, offices, and in the service industry. The workspace is significantly important to do the work, and it should not be a stress buster. Likewise, ergonomics is used for the kitchen as it plans its effective layout.

Middlesworth et al. [5] In the principle of ergonomics, the human factors are related to the interaction between human beings and other things found in the system. This focuses on the welfare of human beings with optimization.

Ghasemi et al. [6] If the model of the kitchen is not based on the principles of ergonomics, then the knowledge of the workers in the kitchen cannot give success.

Rose et al. [7] The environment of the kitchen should be fit for the housewives/cooks so that they will not get tired easily. Normal fatigue can be reduced by rest and sleep. Therefore work-related fatigue should be reduced to minimize health issues.

Nagai et al. [8] Personal factors (physical, mental, behavioral, no time to relax) and organizational factors (work attitude, performance level, injuries) are observed.

Stone et al. [9] The purpose of this research was to examine the impact of knife features on consumers' ability to cut vegetables. There are numerous factors to consider when purchasing a knife, and various opinions on what makes for the ideal knife. In this research, an affordance instrument was developed, produced, and evaluated. The purpose of this attachment is to make a pinch grip on the knife more accessible. This grip improves cutting efficiency and precision by facilitating a more natural position of the wrist and forearm. Sixteen people were put through the tool's tests to make sure it wouldn't need more muscle activity, take longer, be more uncomfortable, or result in worse slice performance. In the end, utilising the affordance tool as planned did not compromise cutting performance when compared to conventional knife use.

Shete et al. [10] In the kitchen everything should be available easily, and fewer numbers of movements should be there.

Suhas et al. [11] Muscle stress should be reduced by minimizing the physical movements i.e. chopping the vegetables, holding down the hand for cooking the food, bending down to find or arrange the utensils on the shelves, lifting high to find the things.

Middlesworth et al. [5] The kitchen plays a significant role when we consider the health of housewives'/cooks'. Many of them have reported muscular disorders i.e. lower and upper neck pain. Therefore, the kitchen environment should be based on ergonomic principles.

Maguire et al. [12] Most people have trouble reaching the upper shelves without slouching or expanding their pelvis, reaching the bottom shelves without leaning down, and raising to reach the top shelves. Keeping the housewives and chefs from being distracted while they work, most tasks in the kitchen are performed with their backs to the walls.

Bonenberg et al. [13] There are six kinds of spatial layouts for the kitchen i.e. single-row, L-shaped, double-row, U-shaped, and kitchen furniture with an island or a semi-island. The senior or physically challenged persons should choose the best layout for their kitchen. All these layouts are good for the disabled persons sitting on the wheelchairs for their moving up in the kitchen. Figure 3 shows the kitchen adjusted to persons moving in a wheelchair.

Eiler et al. [14] In the 1940s, the research at the University of Illinois School of Architecture proposed that the construction of the standardized kitchen is cost-effective

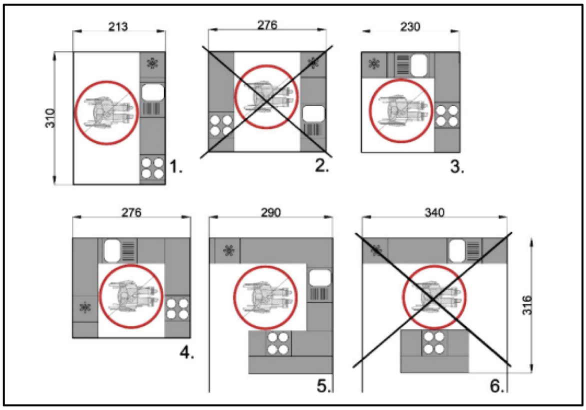


Fig. 3. The kitchen adjusted to persons moving in a wheelchair [13]

and there should be the work triangle that is an imaginary line going from the sink to the cooking stove to the refrigerator and then back to the sink.

Busch et al. [15] gave postural optimization theory. This works on safety and acceptability that combine a postural techniques, human model, and movement tracking system. They assess the static postures.

Ferraguti et al. [16] give a new principle of ergonomics in which the robot helps the housewives'/cooks' by adjusting the position of the goods at convenient places so that the posture will be good, and any muscular problem should not be in the body.

Williams et al. [17] If the kitchen follows Gally shaped model, then the kitchen triangle proves outdated by the architecture because it is hard to set up due to the narrow space in the kitchen. Additionally, concept must support sustainable infrastructure [18, 19].

Evans et al. [20] The new recommendations are based on the consumers' health also to prevent any contamination in their food. "This is a new suggested concept and although in this study we presented data that supports our concept, we acknowledge there are limitations such as (a) sample size (64 households), (b) other factors that could cause cross-contamination events (consumers' level of knowledge, routines and foodborne risk perception), (c) outliers (consumers lacking basic means) and (d) consumers' behavior that can change under observation" [20].

Zhu et al. [21] This research describes the expectations of the future generation for their kitchen. One laboratory experiment was done with the help of 51 participants to show the expectations of the housewives/cooks for their kitchen. The participants sketched a sketch of their future kitchen. The outcomes show a four-factor 17 index instrument for discovering the characteristics of future kitchens i.e. Sustainability, expansibility, interaction, smart technology, and emotional relation. This gives a layout of the comprehensive model of the kitchen. The obtained result showed the important implications for designing smart kitchens.

Earlier studies found that the location of the sink in the kitchen layout had a significant impact on consumers' hygiene practices. Placing sink near to food cooking area would improve hygiene practices [22]. In recent years, the concept of work triangle is used as parameter for planning efficient kitchen [23–26].

Gerhard et al. [27] have worked on use of Mathematical Modeling, Statistical Learning, Optimization and Data Science, Heuristics and Artificial Intelligence (AI) on different disciplines of engineering and real-life problems.

3 Conclusion

The major cause of muscle pain during kitchen operation is stretching on the pelvis, lifting to access the top shelves, and bending down for the lower shelves. The review presented different strategies on improving design of kitchen using different spatial layouts. Various theories are also postulated that provides detail on safety and acceptability that combine a postural techniques, human model, and movement tracking system. The future scope of ergonomic kitchen design is vast, and with the integration of technology, sustainability, and inclusivity, kitchen designers can create functional and comfortable spaces that meet the needs of the users. Kitchen designers can implement universal design principles to make kitchens more inclusive for everyone i.e. person with disabilities. Future kitchens can be designed with sustainable materials that are safe and easy to clean, reducing the need for harmful chemicals.

References

1. Arbetsmiljöverket, Statistiskacentralbyrån: Arbetsmiljön 2003 (The Work Environment 2003), Sverigesofficiellastatistik. Statistiskameddelanden AM 68 SM 0401 (2004)
2. Perkiö-Mäkelä, M., et al.: Work and Health Survey 2006 (in Finnish). Statistics. Finnish Institute of Occupational Health, Helsinki (2006)
3. Haukka, E., Leino-Arjas, P., Solovieva, S., Ranta, R., Viikari-Juntura, E., Riihimaki, H.: Co-occurrence of musculoskeletal pain among female kitchen workers. Int. Arch. Occup. Environ. Health **80**, 141–148 (2006)
4. Mahajan, S., Patwardhan, S.: A study of ergonomic approach to kitchen work centers. Asian J. Home Sci. **10**(2), 371–374 (2015)
5. Middlesworth, M.: Ergonomics 101: the definition, domains, and applications of ergonomics (2020). Retrieved from https://ergo-plus.com/ergonomics-definition-domains-applications/
6. Ghasemi, M., Menshadi, D.F., Soltani, A., Sheykhzadeh, H.: Study of knowledge and performance of automobile workers about of health care for back bone. J. Health Sch. Health Res. Inst. **3**(9), 53–60 (2005)
7. Rose, D., et al.: Associations of fatigue to work-related stress, mental and physical health in an employed community sample. BMC Psychiatry **17**(1), 167 (2017)
8. Nagai, M., et al.: Effects of fatigue on immune function in nurses performing shift work. J. Occup. Health 1107180194 (2011)
9. Stone, R.T., Janusz, O., Schnieders, T.M.: Ergonomic analysis of an affordance tool for modern kitchen knives. In: Proceedings of the Human Factors and Ergonomics Society 2019 Annual Meeting (2019)
10. Shete, K., Pandve, H., Puntambekar, T.: Role of ergonomics in kitchen related back problems. J. Ergon. **5**(3), 1000–1141 (2015)
11. Suhas, M., Prajapathi, P., Jolhe, D., Lakhe, R.: A study of Indian kitchen from ergonomic consideration. Ind. Eng. J. **12** (2019)

12. Maguire, M., et al.: Kitchen living in later life: exploring ergonomic problems, coping strategies and design solutions. Int. J. Des. **8**(1), 73–91 (2014)
13. Bonenberg, A.: Designing a functional layout of a kitchen for persons with disabilities – concept of optimal access points
14. Eiler, L.: Kitchen layout 101: the work triangle & zones. CRD Des. Build. (2019). https://www.crddesignbuild.com/blog/kitchen-layout-101-the-work-triangle-and-zones
15. Busch, B., Guilherme, M., Yoan, M., Marie, D., Manuel, L.: Postural optimization for an ergonomic human-robot interaction. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, Canada (2017)
16. Ferraguti, F., Villa, R., Talignani Landi, C., Zanchettin, A.M., Rocco, P., Secchi, C.: A unified architecture for physical and ergonomic human–robot collaboration. Robotica **38**(4), 669–683 (2020)
17. Williams, T.: Is the kitchen work triangle an outdated design rule? Freshome (2020). https://www.mymove.com/home-inspiration/kitchen/kitchen-work-triangle/
18. Kour, K., et al.: Smart-hydroponic-based framework for saffron cultivation: a precision smart agriculture perspective. Sustainability **14**(3), 1120 (2022). https://doi.org/10.3390/su14031120
19. Singh, D., Thakur, A., Singh, M., Sandhu, A.: IoT implementation strategies amid COVID-19 pandemic. Int. J. Comput. Appl. Technol. **65**(4), 389–398 (2021)
20. Evans, E.W., Redmond, E.C.: Behavioral observation and microbiological analysis of older adult consumers' cross-contamination practices in a model domestic kitchen. J. Food Prot. **81**(4), 569–581 (2018). https://doi.org/10.4315/0362-028X.JFP-17-37
21. Zhu, S., Xu, Z., Dong, Y.: What will the future kitchen look like? An exploratory laboratory study of the future expectations of Chinese Generation Z. Int. J. Ind. Ergon. **87**, 103259 (2022)
22. Mihalache, O.A., Borda, D., Dumitras, L.: Kitchen layouts and consumers' food hygiene practices: ergonomics versus safety. Food Control **131**, 108433 (2022). https://doi.org/10.1016/j.foodcont.2021.108433
23. Adams, C.: What Is the Kitchen Triangle? ThoughCo. (2018). https://www.thoughtco.com/kitchen-work-triangle-1206604
24. Wallender, L.: Use the kitchen triangle for space planning. The Spruce (2020). https://www.thespruce.com/classic-kitchen-triangle-1822213. Accessed 12 Mar 2021
25. Pehkonen, I., et al.: Evaluation of a participatory ergonomic intervention process in kitchen work. Appl. Ergon. **40**(1), 115–123 (2009). https://doi.org/10.1016/j.apergo.2008.01.006
26. Hagan, J.A., Kwofie, A.S., Baissie, F.: Assessing the effect of kitchen layout on employee's productivity. IJRDO J. Appl. Manag. Sci. **3**(1), 36–45 (2017)
27. Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022). https://link.springer.com/book/10.1007/978-3-031-19958-5

Tangible Factors Affecting Tourist Satisfaction and Revisit Intention in Gandaki Province, Nepal

Pradeep Sapkota[1,2](✉), Kamal Kant Vashisth[1], Deepesh Ranabhat[2], and Shanti Devi Chhetri[2]

[1] Faculty of Management Sciences, Shoolini University, Bajhol, Himachal Pradesh 173229, India
ursparu061@gmail.com
[2] Faculty of Management Studies, Pokhara University, Pokhara, Kaski, Nepal

Abstract. Tourist satisfaction is the extent to which tourists are pleased with their overall travel experience. Various tangible and intangible factors associated with tourism affects tourist satisfaction and revisit intention. This study examines the tangible factors affecting tourist satisfaction and revisit intention among the international tourist visiting Gandaki province, Nepal. A total of 400 tourists were taken as sample and data was collected using self-administrative questionnaire. Data analysis was conducted using IBM SPSS. Mean score analysis, exploratory factor analysis, and confirmatory factor analysis were applied in this study. The study findings show that price, infrastructure, and safety and security are the major tangible factor related to tourist satisfaction and revisit intention. The study concludes that the service provider should focus on reasonable price, well-developed infrastructure, and proper safety and security of tourist product and services.

Keywords: Price · Infrastructure · Safety and security · Tourist satisfaction · Revisit intention

1 Introduction

Tourist satisfaction is a crucial aspect of the success of the tourism industry. It refers to the level of contentment that tourists experience during and after their trip. The satisfaction of tourists has been widely studied, as it has significant implications for the tourism industry and destination management. It is important for the tourism industry to understand the factors affecting tourist satisfaction, as it can help to improve the experience of tourists and build a formidable reputation for the destination. A successful judgment of tourist's satisfaction encourages their facilities diversification, enhances their retention, increases competitiveness and positive words of mouth to the others. Tourist's satisfaction with their attraction sites is a result of many aspects, such as their perception of tourism facilities and services, experiences as well as their expectations before and during their visits. Understanding what motivates tourist satisfaction is one of the most important areas of study for the tourism sector [1], as satisfied tourists likely to revisit the place again and recommend others to visit the place [2].

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 410–419, 2024.
https://doi.org/10.1007/978-3-031-50151-7_39

Several researchers from around the world have identified various factors that affect the tourist's satisfaction based on the quality of goods and services that also affect the revisit attitudes. Some of these factors that affect tourists' satisfaction are time, benefits, price, effort, perceived value, cultural differences, destination image and experience quality [3, 4]. However, according to the available literature, there is no study on the tangible factors affecting tourist satisfaction and revisit attitudes covering Gandaki Province, Nepal. This study seeks to fill the gap in the quality of products and services and tourists' satisfaction in a developing country (Nepal) particularly Gandaki regions of this country and it provides useful knowledge for different tourism sectors managers to improve their products and services for the tourist's satisfaction in infrastructure, price, safety and security and other related categories. Therefore, the focus of this study is to establish and evaluate the major tangible factors that influence the satisfaction of tourists and to examine the relationship between the overall quality of the products and services with tourist's expectations in Gandaki Province of Nepal. Moreover, it also focuses on establishing and evaluating the factors that affect tourist's revisit attitude to Gandaki Province of Nepal. Furthermore, it fills the gap for the lack of additional literature regarding the said factors and raises information about these factors.

2 Literature Review

Tourist satisfaction is influenced by a range of factors, including both tangible and intangible factors [5]. Anyone responsible for developing and marketing tourism-related products should be aware of these elements. The model proposed [6] on service blueprint is the most popular approach for getting insights into tangible and intangible factors of tourism products. A various factors affect tourist satisfaction including the tangible and intangible factors such as the quality of facilities, price of product, amenities, and infrastructure at the destination (tangibles), local culture and customs, emotional experiences, perceived value (intangibles). Tangible factors can play a crucial role in shaping tourists' experiences and perceptions of a destination [7] and are an essential consideration for destination managers and marketers in developing and promoting their tourism products.

Price is a significant factor that can influence tourist satisfaction. Tourists who perceive the price of their travel to be reasonable are more likely to be satisfied with their experience and to recommend the destination to others [8]. The impact of price on satisfaction differ depending on the type of tourist, transportation, food and the type of hotel [9]. Price can also play a role in determining repeat visitation, with satisfied tourists more likely to revisit a destination in the future. Infrastructure is a crucial factor that can impact tourist satisfaction and influence their decision to return to a destination. Various has shown that high-quality infrastructure is essential for improving tourist satisfaction and loyalty [10]. Applied structural equation modelling technique with sample size of 336 respondents. The findings of the study suggest that visitor happiness is positively impacted by the quality of infrastructure available in the tourism destination. Similarly, [11] conducted a study to find out the reason for revisit intention in Ho Chi Minh City. Result suggested that infrastructure facilities like condition of road, quality of hotel, facilities of drinking water were most important along with natural and cultural elements. Safety and security are essential for offering high-quality tourism and should be the top priority of each tourism destinations. The reputation of tourism destination and visitor satisfaction depend on the issue of safety and security [12]. A study conducted by [13] taking 347 migrant visitors on three selected beach of Ghana suggested that safety and security is crucial factor tourist arrival and Ghana's tourism development. Overall, the studies shows that price, infrastructure and safety and security are the key factor to consider in destination management, and its provision and maintenance can significantly affect the level of satisfaction and revisit intentions.

3 Research Methodology

This study follows cross-sectional research design. Those tourists who visited Gandaki Province, Nepal were the population units for this study, and 400 tourists were taken for data collection. The researchers collected the necessary data using a self-administered questionnaire from January to December 2022 from different destinations of Gandaki province. The data included demographic information and Likert Scale items to assess tourist satisfaction. IBM SPSS was used to analyze the data. In this research, means, standard deviations (SD), exploratory factor analysis (EFA), and confirmatory factor analysis (CFA) were used.

4 Data Analysis and Results

This includes socio-demographic characteristics, mean score analysis of tourist satisfaction, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA).

4.1 Socio-demographic Characteristics

Table 1 presents the socio-demographic characteristics including gender, age, marital status education, monthly income, expenses per visit, times of visit, and average length of stay of respondents.

Table 1 reveals that more than half of the respondents (54.2%) were males and married (48.5%). Similarly, larger portion (37.5%) of respondents were from 31 to 40 years of age. Majority (66.6%) of respondents had education up to postgraduate level and above. 50% of respondents have monthly income level more than $2000 whereas, 4.3% of respondents didn't have income. Regarding expenses almost 50% respondents spend more than $750. Similarly, majority (76.8%) respondents were first time visitors whereas, 4.5% have visited more than 3 times. Likewise, more than 71% respondents spend more than 7 days in Gandaki province and just 4.3% respondents spend less than 3 days.

4.2 Opinion Towards Tourist Satisfaction and Revisit Intention

This study used 30 items to measure tourist satisfaction and revisit intention in Gandaki Province. A five-points Likert Scale was used, with 1 indicating strongly disagree and 5 indicating strongly agree. Table 2 shows the mean score of the items.

Table 1. Socio-demographic characteristics of respondents

Variable	Categories	Frequency	Percent
Gender	Male	217	54.3
	Female	183	45.8
Age of respondents	20 or below	10	2.5
	21–30	85	21.3
	31–40	150	37.5
	41–50	80	20.0
	51–60	61	15.3
	Above 60	14	3.5
Marital status	Married	194	48.5
	Unmarried	182	45.5
	Others	24	6.0
Education	Secondary education	35	8.8
	Undergraduate	99	24.8
	Postgraduate	237	59.3
	PhD and above	29	7.3
Monthly income ($)	No income	17	4.3
	Up to 500	13	3.3
	501–1000	35	8.8
	1001–1500	40	10.0
	1501–2000	95	23.8
	More than 2000	200	50.0
Expenses per visit ($)	Up to 250	32	8.0
	251–500	149	37.3
	501–750	120	30.0
	751–1000	79	19.8
	More than 1000	20	5.0
Times of visit in Gandaki	First time	307	76.8
	Second time	60	15.0
	Third time	15	3.8
	More than third time	18	4.5
Length of stay in Gandaki	1–2 days	17	4.3
	3–4 days	97	24.3

(*continued*)

Table 1. (*continued*)

Variable	Categories	Frequency	Percent
	5–7 days	139	34.8
	More than 7 days	147	36.8
Total		400	100.0

Table 2 presents all items used to measure tourist satisfaction and revisit intention in Gandaki province, Nepal. It consists of 3 major tangible factors, infrastructure, price and safety and security. The mean value more than the average of 3 indicates that tourists are satisfied with the factors. The first factor infrastructure has mean score 2.5–3.26. Respondents have positive response on the item quality hotels and lodges are available in the destination with mean score of 3.26 whereas, the mean score of less than 3 denotes that there is no proper road connectivity to tourism sites and no well-developed market. Similarly, mean score of price ranges from 3.65 to 4.17, denotes that respondents are satisfied with price. Likewise, mean score for safety and security range from 3.8 to 4.37. The item satisfaction has mean score ranging from 4.0 to 4.37 indicating that majority of respondents agreed that they are satisfied with the destination. Similarly, the items included in revisit intention factor has mean score range from 3.77 to 3.95, which shows respondents positive response towards revisit intention.

4.3 Exploratory Factor Analysis

An exploratory factor analysis (EFA) was performed with 30 items to identify the main tangible factors associated with tourist satisfaction and revisit intention. The communalities of four items – INF1, PRI4, SAS1, and SAS2 were found below than the minimum desirable limit of 0.5 and four items – INF2, SAT1, SAT4, and REV2 were cross loaded on more than one factors, so these items were removed. Finally, factor solution was achieved with 22 items. The communalities of 22 items were found more than 0.50, ranges from 0.545 to 0.827. Similarly, the suitability of EFA was checked with Kaiser–Meyer–Olkin (KMO) and Bartlett's test of sphericity. The KMO value 0.848 is more than the minimum required value of 0.6 indicates the sample size is appropriate [14]. Likewise, Bartlett's test of sphericity is significant at 0.001 indicates that there is high association between the items which is appropriate for factor analysis [15].

The factor analysis result was determined by using varimax rotation and eigenvalues larger than one, which were found in five factors. The first factor comprises 6 items (SAT2, SAT3, SAT5, SAT6, SAT7, and SAT8) explained 16.33% of variance, the second factor comprises 5 items (PRI1, PRI2, PRI3, PRI5, and PRI6) explained 13.846% of variance, the third factor contains four items (REV1, REV3, REV4, and REV5) explained 13.278% of variance, the fourth factor contains four items (SAS3, SAS4, SAS5, and SAS6) explained 12.00% of variance, and the fifth factor contains three items (INF3, INF4, and INF5) explained 8.764% of variance. These factors were named tourist satisfaction, price, revisit intention, safety and security, and infrastructure respectively that explained total 64.221% of the variance. The EFA result is presented in Table 3.

Table 2. Mean score related to tourist satisfaction and revisit intention

Items code	Items	Mean	SD
INF1	Well-developed market	2.77	1.041
INF2	Proper road connectivity	2.50	1.050
INF3	Quality way side amenities	3.06	1.012
INF4	Availability of vehicles	3.25	0.913
INF5	Quality hotels and lodge	3.26	1.023
PRI1	Price of food and beverage	4.17	0.803
PRI2	Price of touristic products	3.92	0.903
PRI3	Price for entry permit	3.65	1.075
PRI4	Price for basic medical facilities	3.80	0.855
PRI5	Price for tour package	3.91	0.808
PRI6	Price for adventurous activities	4.03	0.792
SAS1	Strict law and order	3.80	0.874
SAS2	Feeling of safe and secure	4.30	0.709
SAS3	Absence of nuisance	3.91	1.029
SAS4	No misbehave by locals	4.34	0.782
SAS5	No fear of criminal activities	4.37	0.761
SAS6	No fraud by vehicle staff	3.95	0.994
SAT1	Destination exceed expectation	4.00	0.784
SAT2	Enjoyed the destination	4.37	0.666
SAT3	Good experience	4.42	0.648
SAT4	Offered value of money	4.09	0.815
SAT5	Give lot of excitement and pleasure	4.26	0.715
SAT6	Acceptable standard of quality	4.18	0.707
SAT7	Satisfaction with the service	4.25	0.712
SAT8	Overall satisfaction	4.34	0.631
REV1	Visit destination in near future	3.94	0.984
REV2	Recommend to others	4.54	0.624
REV3	Not enough to explore in one visit	3.95	0.933
REV4	Bring family members and friends	3.77	0.957
REV5	Learn new things by visiting again	3.92	0.910

(Where, N = 400, 1 denotes strongly disagree and 5 denotes strongly agree)

Table 3. EFA result

Items	Loadings	% of variance	Cumulative %	Cronbach's alpha
SAT2	0.782	16.333	16.333	0.860
SAT3	0.784			
SAT5	0.651			
SAT6	0.589			
SAT7	0.688			
SAT8	0.806			
PRI1	0.709	13.846	30.179	0.821
PRI2	0.834			
PRI3	0.703			
PRI5	0.752			
PRI6	0.753			
REV1	0.803	13.278	43.457	0.861
REV3	0.741			
REV4	0.751			
REV5	0.860			
SAS3	0.761	12.000	55.457	0.785
SAS4	0.741			
SAS5	0.788			
SAS6	0.780			
INF3	0.802	8.764	64.221	0.702
INF4	0.712			
INF5	0.793			

4.4 Confirmatory Factor Analysis (CFA)

After getting the results of EFA, confirmatory factor analysis (CFA) was run to confirm the result. Initially, CFA was carried on five factors extracted from EFA with 22 items. But, only 20 items were loaded in CFA. Figure 1 shows the CFA model.

Table 4 presents the results of measurement model. The CMIN/DF value (2.353), GFI value (0.914), CFI value (0.935), and RMSEA value (0.058) meet the required criteria as given in the table. Likewise, NFI value (0.893) is very near to the required criteria. This provides strong evidence for fitness of measurement model.

4.5 Reliability and Validity Analysis

Cronbach Alpha and Composite Reliability (CR) were used to assess the construct reliability with a minimum threshold value of 0.70 [15]. Likewise, the Average Variance

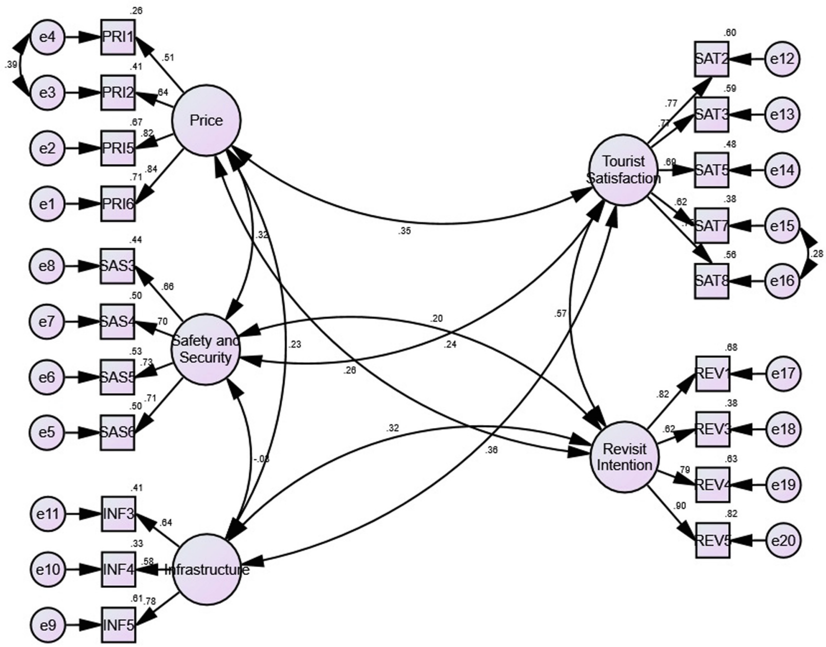


Fig. 1. Result of CFA

Table 4. Summary of model fit indices

Fit indices	Criteria	Calculate value	Remarks
CMIN/DF	< 3	2.353	Well fitted
GFI	0.9 or above	0.914	Well fitted
NFI	0.9 or above	0.893	Moderately fitted
CFI	0.9 or above	0.935	Well fitted
RMSEA	< 0.08	0.058	Well fitted

Extracted (AVE) statistic was used to establish convergent validity. The desirable value of AVE is 0.50 or more, however, AVE of 0.40 or more is also acceptable when CR is more than 0.70. Fornell and Larcker Criteria was performed to establish the discriminant validity. It is achieved when the square root of AVE of construct is greater than the correlation coefficient between of construct with all other constructs [16].

Table 5 displays the Cronbach Alpha, Composite Reliability (CR), and AVE findings. Cronbach's Alpha and CR are both greater than the required threshold of 0.70 in this research which confirm the reliability. Similarly, the AVE for price, tourist satisfaction, and revisit intention exceeds the minimum value of > 0.5 and AVE for safety and security, and infrastructure are more than 0.40 which is also acceptable when CR is greater than 0.70. This provides strong evidence for construct validity.

Table 5. Construct reliability and construct validity

Construct	Cronbach's alpha	Composite reliability	AVE
Price	0.817	0.802	0.512
Safety and security	0.785	0.794	0.491
Infrastructure	0.702	0.708	0.451
Tourist satisfaction	0.845	0.844	0.521
Revisit intention	0.861	0.869	0.627

Table 6. Discriminant validity - Fornell Larcker's criteria

	Pr	S&S	Infr	TS	RI
Price (Pr)	***0.716***				
Safety and security (S&S)	0.316	***0.701***			
Infrastructure (Infr)	0.228	− 0.030	***0.672***		
Tourist satisfaction (TS)	0.347	0.241	0.357	***0.722***	
Revisit intention (RI)	0.257	0.199	0.325	0.569	***0.792***

In Table 6, the diagonal value represents the square root of AVE and other values represents the correlation coefficients between the constructs. The square root of AVE of price (0.716), safety and security (0.701), infrastructure (0.672), tourist satisfaction (0.722), and revisit intention (0.792) are greater than the correlation value. Therefore, discriminant validity was established.

Overall, the results of CFA, reliability and validity tests confirm that price, safety and security, and infrastructure are important tangible factors related to tourist satisfaction and revisit intention.

5 Conclusion

Tourist satisfaction is a crucial aspect of the tourism industry, as it directly impacts the success of destinations, attractions, and service providers. Meeting and exceeding tourist expectations requires careful attention to the various factors that contribute to their overall experience. There are several factors which can affect the arrival of tourists and their repeat visit intentions.

The study was conducted to find out the tangible factors related to tourist satisfaction and revisit intention in Gandaki province, Nepal. This study found that tourist is satisfied with the reasonable price for food and beverage, touristic products, tour package, basic medical facilities, availability of tourist vehicle, quality of way side amenities, hotels and lodge, and safety and security in Gandaki province. This study also found that price, safety and security, and infrastructure are the major tangible factors related to tourist satisfaction and revisit intention in the study area. Likewise, it was observed that price was the most important factor whose items covered 13.846% of total variance followed by safety and security (12%) and infrastructure (8.764%) of total variance. This study concludes that the service provider should focus on reasonable price of tourist products and service, well developed infrastructure and proper safety and security of the tourist for satisfaction and revisit. This study only extracted the tangible factors related to satisfaction and revisit intention. So, further research could be done by examining the impact of tangibles factors as well as intangible factors affecting tourist satisfaction and revisit intention.

References

1. Petrick, J.F., Morais, D.D., Norman, W.C.: An examination of the determinants of entertainment vacationers' intentions to revisit. J. Travel Res. **40**(1), 41–48 (2001). https://doi.org/10.1177/004728750104000106
2. Su, A.Y.L.: Customer satisfaction measurement practice in Taiwan hotels. Int. J. Hosp. Manag. **23**(4), 397–408 (2004). https://doi.org/10.1016/j.ijhm.2004.02.002
3. Yoon, Y., Uysal, M.: An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. Tour. Manag. **26**(1), 45–56 (2005). https://doi.org/10.1016/j.tourman.2003.08.016
4. Huang, Q., Satchabut, T.: Factors affecting foreign tourists' satisfaction and willingness to pay more for visiting Bangkok. UTCC Int. J. Bus. Econ. **11**(2), 1–22 (2016)
5. Hosany, S., Gilbert, D.: Measuring tourists' emotional experiences toward hedonic holiday destinations. J. Travel Res. **49**(4), 513–526 (2010). https://doi.org/10.1177/0047287509349267
6. Shostack, G.L.: How to design a service. Eur. J. Mark. **16**(1), 49–63 (1982). https://doi.org/10.1108/EUM0000000004799
7. Suanmali, S.: Factors affecting tourist satisfaction: an empirical study in the northern part of Thailand. SHS Web Conf. **12**, 01027 (2014). https://doi.org/10.1051/shsconf/20141201027
8. Campo, S., Yagüe, M.J.: Research note: effects of price on tourist satisfaction. Tour. Econ. **14**(3), 657–661 (2008). https://doi.org/10.5367/000000008785633596
9. Ngurah, G., et al.: Factors that influences tourist's satisfaction and its consequences. Eur. J. Bus. Manag. **9**(8), 39–50 (2017). [Online]. Available: www.iiste.org
10. Sugiama, A.G., Oktavia, H.C., Karlina, M.: The effect of tourism infrastructure asset quality on tourist satisfaction: a case on forest tourism in Tasikmalaya Regency. Int. J. Appl. Sci. Tour. Events **6**(1), 65–71 (2022). https://doi.org/10.31940/ijaste.v6i1.65-71
11. Tien, N.H., Dung, N.T.H., Trang, T.T.T., Hien, V.T., Phuong, B.T.N.: Factor affecting tourists' return intention. A case of Binh Quoi Village in Ho Chi Minh City. PalArch's J. Archaeol. Egypt Egyptol. **18**(09), 493–507 (2021). [Online]. Available: https://archives.palarch.nl/index.php/jae/article/view/9334
12. Christopher Holloway, C.H.J.: Front Cover Image for the Business of Tourism the Business of Tourism
13. Preko, A.: Safety and security concerns at the beach: views of migrant visitors in Ghana. Tour. Hosp. Res. **21**(1), 73–85 (2021). https://doi.org/10.1177/1467358420954504
14. Kaiser, H.F., Ford, H.: An index of factorial simplicity. Psychometrika **39**(1), 31–36 (1984)
15. Hair, J.F., Ringle, C.M., Sarstedt, M.: PLS-SEM: indeed a silver bullet. J. Mark. Theory Pract. **19**(2), 139–152 (2011). https://doi.org/10.2753/MTP1069-6679190202
16. Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. J. Acad. Mark. Sci. **43**(1), 115–135 (2014). https://doi.org/10.1007/s11747-014-0403-8

Determining the Optimal Formulation of Cyanoacrylate Nanocomposition to Ensure Strain-Free Fixation of Workpieces on CNC Mills

A. S. Kononenko[1](✉), R. R. Khabbatullin[1], I. Yu. Ignatkin[2], and A. Yu. Alipichev[2]

[1] Bauman Moscow State Technical University, Moscow 105005, Russia
as-kononenko@yandex.ru
[2] Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow 127550, Russia

Abstract. The paper considers the possibility of upgrading the accuracy of machining body parts by strain-free fixing of workpieces with cyanoacrylate adhesives. The authors consider the ways to increase the durability of cyanoacrylate polymeric compositions by introducing nanoparticles in their matrix. The optimum proportions of polymer and nanofillers have been theoretically proved to increase the elasticity modulus of the composition. The paper presents the results of experimental studies confirming the positive effect of nanoparticles on the strength properties of the cyanoacrylate polymers.

Keywords: Cyanoacrylate polymer · Nanocomposition · Nanosize powders · Silicon oxide · Aluminum oxide

1 Introduction

It is well known that the quality parameters of engineering products are mainly dependent on the accuracy of their main components. Body parts belong to the most critical and complex engineering products, which account for the vast majority of manufacturing output. In turn, the quality indicators of the assembly are directly dependent on the accuracy indicators of the body parts, more than 20% of them are considered to be of high-precision. The highest requirements are imposed on the form accuracy of main holes and relative spatial position of their axes [1].

Most body parts are machined on multifunctional CNC mills performing operations during one setup with several tool transitions and variable workpiece positioning. In this case, different-shape surfaces of the workpiece are machined in the same fixture to satisfy the requirements of its accuracy, immobility or quasi-static relative motion during machining. In this context, it is rather urgent to determine the method of fixing the machined workpiece [2].

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 420–427, 2024.
https://doi.org/10.1007/978-3-031-50151-7_40

When choosing a method of workpiece fixing, it is necessary to ensure a well-defined position of the workpiece in relation to the cutting tool. Fixing can be complicated by the fact that many workpieces can be deformed while machining. However, decreasing fixing forces can cause the machined workpiece to move under the effect of cutting forces [3].

The factors influencing the total machining error have a reciprocal influence, so machining errors occurring at the initial stage are inherited until the end of the process. The inheriting component in the accuracy of the main dimensions ranges between 40 and 50% of the machining quality index value [2].

To achieve high accuracy and quality of machining, in addition to solving complex technological problems, it is necessary to consider the influence of elastic deformations of body parts on changes in shape and spatial position of their surfaces, arising at the initial stages of machining - during fixing on the mills.

For single and small batch production, CNC mills with a rotary turntable capable of precise positioning offer an optimum solution. The method of workpiece fixing is of decisive importance. The fixture must not hinder access to the machined surfaces and must provide the opportunity of machining all critical elements in a single setup.

The main purpose of the fixing device on a CNC mill is to prevent the workpiece from moving freely in the X, Y, Z axes and rotating around these axes, respectively. This can be achieved by fixing one of the planes.

The method of fastening workpieces to magnetic plates is now widely used in Russia. Its main idea is to convert the input energy of the magnetic field into the mechanical energy of fastening. However, it is impossible to fastening non-magnetic materials; moreover, pre-treatment is necessary to ensure the tight adjoining of the workpiece to the magnetic machine tool plane. Another method for fastening workpieces is the use of vacuum plates. They produce vacuum in suction cups that are in direct contact with the workpiece. But this method is difficult to use in single-piece and small-scale production (each part requires a change in the vacuum plate design). Besides fixing the workpiece on a vacuum or magnet plate, it is also possible to prevent it from moving without being distorted by gluing it to a substrate, which is a plate mounted in a machine vice on the worktable of the CNC mill [4].

The CNC machining is accompanied by high forces and vibrations that are absorbed by the machined material. Thus, on the one hand, a high adhesive strength is required for reliable fixing of the workpiece, and, on the other hand, the adhesive bond should be easily destroyed without deforming the workpiece after machining [5].

2 Materials and Methods

To fix the workpiece on the mill table, it is advisable to use polymeric materials. Their available range allows selecting the optimum adhesive composition for the specific fixing case depending on the materials to be joined and the machining conditions. Since machining costs are largely dependent on the mill running time, it is advisable to use polymeric compositions with a short time interval for gaining strength. Cyanoacrylate adhesives are among these fast curing polymers.

In Russia, adhesive compositions of such foreign companies as THREE BOND (Japan), Loctite (USA) [6], OOO Tekhnobazis and JSC NII POLYMEROV (Russia) are widely used [7].

A number of sources [8–10] state that the advantages of using polymers include their versatility and ease of application, high penetrating ability, due to which the adhesive composition is uniformly distributed over the whole bonding area, which significantly reduces the degree of possible deformations and distortions resulting from the contact of two surfaces. According to Loctite, most plastic materials break down at temperatures above 150 °C. Accordingly, the substrate can be separated from the machined workpiece simultaneously as the workpiece is heat treated.

The disadvantages of using polymeric materials include low elastic properties, strength characteristics, and insufficient resistance to vibration loads, which limits their use for adhesion of workpieces during machining.

However, by modifying polymer compositions with nanosize fillers, we can significantly improve the mechanical properties of polymers [9, 10]. The effect of fillers on the properties of the initial composition depends on the particle size, concentration, chemical nature of the polymer matrix, as well as their type. Carbon nanotubes (hereinafter referred to as CNT), fullerenes and metal oxide nanosize powders are the most commonly used modifiers [10, 11].

When nanosize particles are added to a polymer, the particles can actively participate in the mixing process by forming additional units or centers of structure formation, around which oriented layers of a polymer matrix with a dense packing of components contained in the nanocomposite are formed. Such tie points in a mass composition have a positive effect on the strength of a nanocomposite [12, 13].

The relation between the theoretical strength σt and the modulus of elasticity E of the polymer material is as follows [14]:

$$\sigma_T \approx 0.1E, \tag{1}$$

where E – elasticity modulus, MPa.

The elasticity modulus of metals is two orders of magnitude higher than that of polymers; hence, the filling of polymers with metal particles increases the elasticity modulus of the resulting composition. To describe the degree of increase in the elasticity modulus of nanocomposites, in [15], we used the Kounteau formula:

$$\frac{E_c}{E_\mathrm{p}} = \left(1 - \varphi_\mathrm{f}^{1/2}\right) + \frac{E_\mathrm{f}}{\frac{\left(1-\varphi_\mathrm{f}^{1/2}\right)}{\varphi_\mathrm{f}^{1/2}} \cdot E_\mathrm{p} + E_\mathrm{f}}, \tag{2}$$

where E_c, E_p, and E_f – the elasticity moduli of the composition, unfilled polymer, and filler, respectively, MPa; φ_f – the filler's volume fraction.

The elasticity modulus of cyanoacrylate polymers is about 1.8 GPa [4], and the elasticity moduli of silicon and aluminum oxides are 100 and 360 GPa, respectively. Therefore, according to formula (2), filling the anaerobic polymer with nanosize silicon and aluminum oxide particles will increase the elasticity modulus and consequently the nanocomposite strength. Figure 1 shows the relationship between the relative increase in elastic modulus E_c/E_p and the filler volume fraction φ_{f} for the nanocomposition (cyanoacrylate polymer - silicon oxide) and Fig. 2 - for the nanocomposition (cyanoacrylate polymer - aluminum oxide).

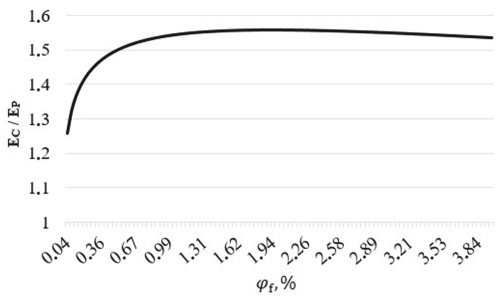


Fig. 1. Relationship between the relative increase of the elastic modulus E_C/E_P and the mass fraction of the filler φ_{f}, for the nanocomposition (cyanoacrylate polymer - silicon oxide)

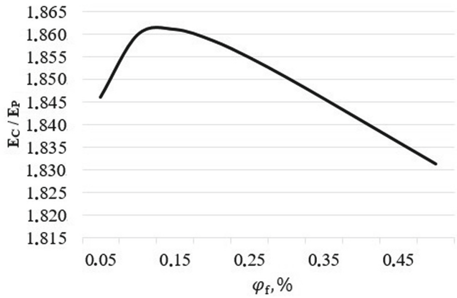


Fig. 2. Relative elastic modulus increase of E_C/E_P as a function of the filler mass fraction φ_{f} for the nanocomposition (cyanoacrylate polymer - aluminum oxide).

The deformation properties of polymer nanocomposites can be enhanced by the fact that, at low volume contents, the nanosize particles move freely in the polymer matrix, which is easily stretchable. In this case, the viscous drag on the movement of filler particles with high adhesion to the polymer matrix may increase the tensile strength properties of the nanocomposite. In contrast, polymers with microsized filler particles are characterized by lower mobility and have higher stresses, which results in lower deformation of the composite [13].

For convenience, let us convert the volume fraction into the mass fraction:

$$\frac{m_f}{m_c} = \frac{\varphi_{\mathrm{f}}}{\rho_f/\rho_c} \tag{3}$$

where ρ_f – the filler density (2650 kg/m^3 for SiO_2 and 3950 kg/m^3 for Al_2O_3); ρ_c – the composite density (we assume that it is equal to that of an anaerobic adhesive: 1050 kg/m^3).

Theoretical analysis has shown that the optimum mass fraction of aluminum oxide in a polymer nanocomposition is about 0.15%, the optimum mass fraction of silicon oxide is about 2.0%.

The adhesion-and-cohesion strength of sealing compounds depends on many factors - the method of polymerization, the chemical nature of the material, the type and concentration of the fillers. Therefore, it is advisable to study the effect on the strength of one factor while keeping the others constant. We decided to estimate the adhesive strength according to the normal destructive stresses (GOST 14760-69). The research method consists in determining the force perpendicular to the bonding plane that causes the composite to detach from the substrate. For the studies, we used samples, each consisting of two cylindrical halves, 25 mm in diameter and 30 mm in length. The working end surfaces of these workpiece halves were machined to a uniform micro-roughening with a roughness Ra = 3.2 μm. After cleaning these surfaces with MS-37 detergent and degreasing with acetone, polymeric cyanoacrylate compositions and nanocompositions based on them were applied and then they were glued. For better mixing of the nanocomposition components, the nanosize powders were pretreated with ultrasound to break up the agglomerations of the particles. After the nanosize particles were introduced into the polymer composition, the components were mechanically mixed for 60 s and the resulting composition was treated with ultrasound at a frequency of 35 kHz for 30 s.

The tests were carried out after complete polymerization of the polymer composition on an R-5 tensile testing machine. The workpiece was loaded at a constant speed of 10 mm/min and gradually increasing load until the polymer adhesive layer was broken. The tensile strength of the polymer composition and nanocomposition based on it was determined from the formula:

$$\sigma_{TS} = \frac{P}{F}, \tag{4}$$

where P – breaking load, N;

F – polymer layer area, m^2.

3 Results

The results of experimental studies of the nanofiller concentration influence on the value of normal destructive stresses of polymer Loctite-496 and nanocompositions based on it are shown in Figs. 3 and 4. Each point in the graphs corresponds to the average of five measurements. The concentration of the nanofillers varied from 0.01 to 2.5% by mass.

4 Discussion

The analysis of the test results showed that the nanomodifying of polymers leads to a change in their strength properties. The addition of SiO_2 nanosize powder in the amount of 2% by mass to the Loctite-496 cyanoacrylate polymer gave a strength increase of 75% (from 12.2 MPa to 21.4 MPa) as compared with the initial values.

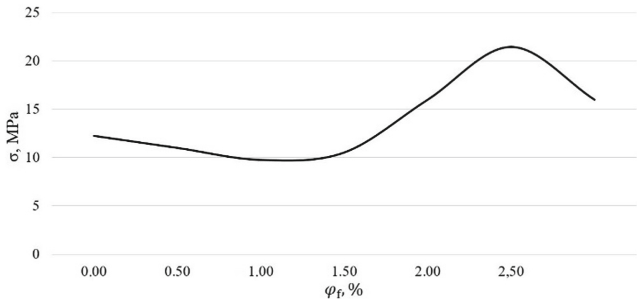


Fig. 3. Relationship between the strength of the composition evaluated by the normal destructive stress σ and the mass fraction of the filler φ_f for nanocomposition (cyanoacrylate polymer - silicon oxide)

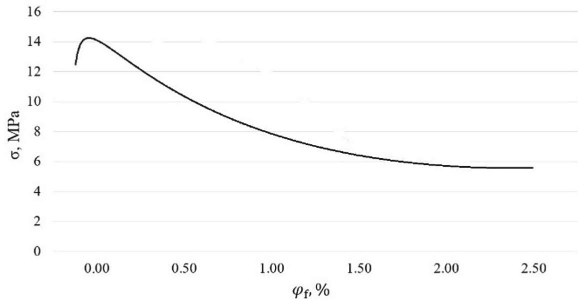


Fig. 4. Relationship between the strength of the composition as estimated from the normal breaking stress σ and the mass fraction of the filler φ_f for nanocomposition (cyanoacrylate polymer - aluminum oxide)

An increase in the Al_2O_3 nanosize powder concentration up to 0.01% raised the normal breaking stress of the Loctite-496 nanocomposition from 12.0 to 14.0 MPa. A further increase in the filler concentration to 2.5% significantly reduced the nanocomposition strength.

The experimental results of determining the nanofiller concentration have a high degree of convergence with the theoretical calculations [16–18].

Also, in the study of the polymer-substrate system, it is important to determine the nature of the bond failure between the components. There are three types of failures: adhesive failure accompanied by complete separation of the adhesive from the substrate; cohesive failure accompanied by a rupture in the adhesive mass; and mixed failure having some signs of adhesive and cohesive failure.

The workpieces were visually inspected after the rupture to determine its nature (Table 1).

The rupture analysis shows that using nanofillers with high surface energy and good wettability considerably improves the adhesion strength component of polymer compositions.

Table 1. Rupture pattern of the polymer layer.

Adhesive sealant	Formulation					
	Cyanoacrylate		Polymer + Al_2O_3		Polymer + SiO_2	
	Rupture type (A - adhesive; C - cohesive)					
	A	C	A	C	A	C
Loctite-496	40	60	25	75	30	70

5 Conclusion

A promising method for the deformation-free fixing of workpieces is the use of polymeric cyanoacrylate compositions to glue them. This method has technological and economic advantages over the methods of fixing workpieces to magnetic and vacuum plates.

The incorporation of nanosize SiO_2 into the polymer matrix of the Loctite-496 cyanoacrylate glue increases the adhesion strength of the composition up to 75%. The best results were obtained by modifying the original polymer with SiO_2 nanosize silicon oxide powder with a mass concentration of 2%.

The modification of cyanoacrylate polymers with SiO_2 and Al_2O nanosize particles increases their strength properties due to the adhesion component of the glue strength.

The results obtained are really relevant for further experimental studies to optimize the mass concentration of nanofillers in cyanoacrylate polymer compositions.

References

1. Poltavets, O.F., Zimin, V.V., Dilanyan, K.R.: Ways to Improve Accuracy of Machining Body Parts on Multipurpose Machines. Review. C-3. Technology and Automation of Manufacture, 90 p. TsNIITEIavtoselmash, Moscow (1990)
2. Vasiliev, A.S.: Total machining error and mutual influence of its components. Izv. Vysshikh Uchebnykh Zaved. Mashinostroenie **2–3**, 89–96 (1999)
3. Berdnikov, L.N., Bezyazychniy, V.F., Krylov, V.N., Bolshakov, E.M., Motrenko, P.D.: Handbook of Milling Machines. In: Bezyazychniy, V.F. (ed.), 272 p. Mashinostroenie, Moscow (2010)
4. Kononenko, A.S., Khabbatullin, R.R.: Theoretical substantiation of the conditions for the applicability of deformationless fixation by means of a polymer glue for workpieces during their mechanical processing on a milling machine with computer numerical control. Polym. Sci. Ser. D **15**(4), 523–528 (2022)
5. Puchin, E.A., Lisunov, E.A., Chepurin, A.V., et al.: Reliability of Technical Systems, 318 p. KoloS, Moscow (2010)
6. Loctite Company: Website. Access mode: https://локтайт.рус/. Accessed 31 Jan 2023
7. JSC "NII Polimerov": Website. Access mode: http://www.nicp.ru/. Accessed 01 Feb 2023
8. Kononenko, A.S., Ignatkin, I.Yu., Drozdov, A.V.: Recovering a reducing-gear shaft neck by reinforced-bush adhesion. Polym. Sci. Ser. D **15**(2), 137–142 (2022)
9. Li, R.I., Psarev, D.N., Kiba, M.R.: Promising nanocomposite based on elastomer F-40 for repairing base members of machines. Polym. Sci. Ser. D Glues Seal. Mater. **12**(2), 128–132 (2019). ISSN 1995-4212

10. Kononenko, A.S., Ignatkin, I.Yu., Drozdov, A.V.: Restoring the neck of a reducing-gear shaft by gluing a reinforced bush. Polym. Sci. Ser. D **15**(4), 529–534 (2022)
11. Panfilov, Y.V., Rodionov, I.A., Ryzhikov, I.A., Baburin, A.S., Moskalev, D.O., Lotkov, E.S.: Ultrathin film deposition for nanoelectronic device manufacturing/vacuum science and equipment. IOP Conf. Ser. Mater. Sci. Eng. **781**, 012021 (2020)
12. Mikhalchenkov, A.M., Tyureva, A.A., Kozarez, I.V., Kononenko, A.S.: The effect of the concentration of components and dispersion of particles of filler of epoxy-sand composite on hardness and its relationship with abrasive wear resistance. Polym. Sci. Ser. D **14**(1), 17–20 (2021)
13. Fanidi, O., Kostryukov, A.A., Shchedrin, A.V., Ignatkin, I.Y.: Comparison of analytical and experimental force in cylindrical workpiece drawing process. Tribol. Ind. **43**(1), 1–11 (2021). https://doi.org/10.24874/ti.1000.11.20.02
14. Narisawa, I.: Durability of Polymer Materials. In: Narisawa, I. Translated from Japanese. Khimiya, Moscow (1987)
15. Kozlov, G.V.: Structure and properties of disperse-filled polymer nanocomposites. UFN **185**(1), 35–64 (2015)
16. Tikhomirov, D., Kuzmichev, A., Rastimeshin, S., et al.: Energy-efficient pasteurizer of liquid products using IR and UV radiation. In: Advances in Intelligent Systems and Computing, vol. 866, pp. 178–186 (2019). https://doi.org/10.1007/978-3-030-00979-3_18
17. Dorokhov, A., Kirsanov, V., Pavkin, D., et al.: Recognition of cow teats using the 3D-ToF camera when milking in the "herringbone" milking parlor. In: Advances in Intelligent Systems and Computing, vol. 1072, pp. 128–137 (2020). https://doi.org/10.1007/978-3-030-33585-4_13
18. Kayabekir, A.E., Nigdeli, S.M., Bekdaş, G., et al.: Adaptive harmony search for cost optimization of reinforced concrete columns. In: Advances in Intelligent Systems and Computing, vol. 371, pp. 35–44 (2021). https://doi.org/10.1007/978-3-030-93247-3_4

Ergonomic Analysis on Kitchen Using Asian Manikin

Amit Bhatia[1], Sandeep Singla[1](✉), and Aarti Bansal[2]

[1] Department of Civil Engineering, RIMT University, Mandi Gobindgarh, Punjab, India
dr.sandeepsinglaz@gmail.com

[2] Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
aarti.bansal@chitkara.edu.in

Abstract. Musculoskeletal injuries are a common problem in kitchens, and kitchen design can play a significant role in preventing them. The physical demands of working in a kitchen, such as standing for long periods, reaching, lifting heavy objects, and performing repetitive tasks, can put a strain on the body, leading to injuries such as strains, sprains, and tendinitis. To reduce the risk of musculoskeletal injuries, kitchen designers should consider ergonomic principles when planning the layout of the kitchen. This includes designing workstations that are at the appropriate height for the tasks being performed, providing adequate space for movement and maneuvering, and placing equipment and supplies within easy reach. The main objective of current research is to carry out the ergonomic analysis on kitchen using Asian manikin. The software used for the analysis is CATIA v5 simulation package. The CAD modelling and ergonomic design analysis is conducted to determine joint stresses, loads acting on different body parts. Using the ergonomic analysis, the kitchen design is analyzed. The analysis results have shown that new kitchen design is ergonomically better than 1st kitchen design.

Keywords: Ergonomics · Kitchen · Modelling · MSD (musculoskeletal disorders)

1 Introduction

Ergonomics, also known as human factors, is the scientific discipline that focuses on understanding and optimizing the interaction between people and their work environment. It involves the study of how people interact with their physical and psychological work environment, with the goal of improving safety, comfort and efficiency. Ergonomics considers various factors such as lighting, noise, temperature, furniture design, equipment design, and workflow. It seeks to optimize the design and layout of workspaces, tools, and equipment to minimize physical strain, reduce the risk of injury, and enhance productivity.

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 428–437, 2024.
https://doi.org/10.1007/978-3-031-50151-7_41

Ergonomics is important in a variety of industries, including healthcare, manufacturing, transportation, and office work. By designing work environments that are tailored to the needs of workers, employers can improve job satisfaction and reduce the risk of work-related injuries and illness. Ergonomics is especially important in the kitchen, where people spend a lot of time performing physical tasks such as chopping, mixing, and cooking. Figure 1 shows the Ideal posture for working at kitchen counter.

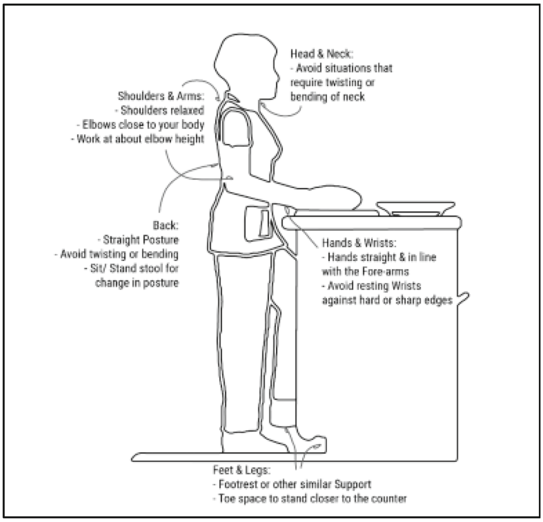


Fig. 1. Ideal posture for working at kitchen counter [1]

- Countertop height: The height of the countertop should be comfortable for the person using it. The ideal height is generally between 36 and 38 in. for someone who is 5′ 6″ to 5′ 8″ tall. However, the countertop height can be adjusted to accommodate the height of the person using.
- Cabinet and drawer height: Cabinets and drawers should be placed at a height that is easy to reach without bending over or reaching too high.
- Lighting: Adequate lighting is essential in the kitchen to prevent eye strain and accidents. Lighting should be bright enough to see everything clearly, but not so bright that it causes glare.
- Appliances: The placement of appliances should be convenient and easy to access. Heavy appliances, such as mixers and food processors, should be kept at a comfortable height and within easy reach.
- Flooring: The flooring in the kitchen should be slip-resistant and easy to clean. It should also be comfortable to stand on for long periods of time.

Poor kitchen design can lead to discomfort, pain, and even injury. One of the most common injuries which occurs due to improper kitchen design is musculoskeletal disorders (MSD). The MSDs affect the musculoskeletal system, including the muscles, tendons, ligaments, nerves, and other soft tissues.

- In the kitchen, MSDs can result from repetitive motions, awkward postures, and excessive force. Some common MSDs that can result from poor kitchen ergonomics include.
- Carpal tunnel syndrome: This condition is caused by compression of the median nerve in the wrist, and can result in pain, tingling, and numbness in the hand and fingers.
- Tendinitis: This is inflammation of the tendons, and can cause pain, swelling, and difficulty moving the affected joint.
- Back pain: Poor posture, heavy lifting, and repetitive bending can all contribute to back pain.
- Neck pain: Looking down for extended periods of time, such as when chopping vegetables, can cause neck pain.
- Shoulder pain: Repetitive motions, such as stirring or lifting heavy pots and pans, can cause shoulder pain.

2 Literature Review

In recent years, there has been a growing interest in kitchen ergonomics for older adults, individuals with dementia, and those with limited mobility. Studies have examined the impact of kitchen design on the independence and quality of life of these individuals, with a focus on incorporating principles of universal design to create a kitchen that is accessible and adaptable for all users.

Earlier research on kitchen ergonomics has demonstrated the importance of designing kitchens that are safe, efficient, and comfortable to use, with a focus on minimizing physical strain and injury, maximizing productivity, and promoting well-being.

Suhas et al. [2] have examined the design of residential kitchens in relation to ergonomics, including the placement of appliances, fixtures, and storage areas. The authors also discussed the importance of accessibility, safety, and ease of use in kitchen design. Arici et al. [3] have focused on kitchen design and equipment for the elderly, with an emphasis on ergonomics and safety. The authors discussed the importance of designing kitchens that minimize physical strain and injury, and provide ease of use and accessibility for elderly users. Selamat et al. [4] have examined the importance of ergonomic considerations in kitchen design, including the placement of appliances and work surfaces, the height of countertops, and the design of cabinets and drawers. The authors emphasized the importance of considering the user's physical characteristics and needs in kitchen design. Pandve et al. [5] have explored consumer needs and habits related to cooking and kitchen design, with a focus on ergonomics and usability. The authors discussed the importance of designing kitchens that are easy to use, efficient, and comfortable for the user. Middlesworth et al. [6] discusses the importance of ergonomics in kitchen design and provides guidelines for creating a functional and ergonomic kitchen space. Kolus et al. [7] provides a comprehensive review of the literature on kitchen ergonomics, including research on physical stress and strain on the body during cooking tasks, as well as ergonomic design principles for kitchen tools and appliances. Lundgren et al. [8] discusses the importance of ergonomic design in the kitchen for reducing physical strain and improving safety during cooking tasks. Antonio et al. [9] compares the ergonomic design features of kitchens in four European countries, highlighting the importance of cultural factors in kitchen design. Mihalache et al. [14] have worked on the relationship between kitchen design, ergonomics, and food safety practices in domestic and commercial kitchens. The study found that kitchen layouts that were more ergonomic and efficient were associated with better food safety practices. Gerhard et al. [19] have worked on use of Mathematical Modeling, Statistical Learning, Optimization and Data Science, Heuristics and Artificial Intelligence (AI) on different disciplines of engineering and real life problems. Özalp et al. [20] have assessed the impact of kitchen ergonomics on the quality of life of the elderly. The study found that a well-designed kitchen can improve the quality of life of the elderly by reducing physical stress and promoting independence.

3 Objectives

Cooking can be a physically demanding task that involves repetitive motions, heavy lifting, and standing for extended periods of time. Without proper ergonomic design, kitchen workers can experience musculoskeletal disorders, such as back pain, shoulder pain, and carpal tunnel syndrome, which can lead to chronic pain and reduced work productivity. An ergonomic analysis of a kitchen is important to ensure that the kitchen is designed in a way that maximizes efficiency, productivity, safety, and comfort. The objective of current research is to conduct the ergonomic analysis on kitchen using Asian manikin. The software used for the analysis is CATIA v5 simulation package. The CAD modelling and ergonomic design analysis is conducted to determine joint stresses, loads acting on different body parts.

4 Methodology

The ergonomic testing is conducted on Asian women manikin with sustainable infrastructure [10–13] using CATIA V5 software. Different operations are conducted which include opening top cupboard, opening bottom cupboard, lightening gas.

In the 1st case top cupboard is opened by the manikin as shown in Fig. 2(a). The opening of cupboard induced loads on different joints, muscles. These include body load compression, axial twist compression, joint shear and abdominal force. For cupboard opening operation, the L4–L5 moment obtained is 12 N m, L4 = L5 compression obtained is 2263 N. The total ground reaction obtained is 633 N. The reaction on left foot is 278 N and reaction on right foot is 355 N. The load on spine is 2263 N. For opening lower cupboard as shown in Fig. 2(b), the L4–L5 moment obtained is 9 N m, L4–L5 compression obtained is 1231 N. The total ground reaction obtained is 613 N. The reaction on left foot is 291 N and reaction on right foot is 322 N. The load on spine is 1231 N.

For turning gas lighter as shown in Fig. 3(a), the L4–L5 moment obtained is 8 N m, L4–L5 compression obtained is 1979 N. The total ground reaction obtained is 613 N.

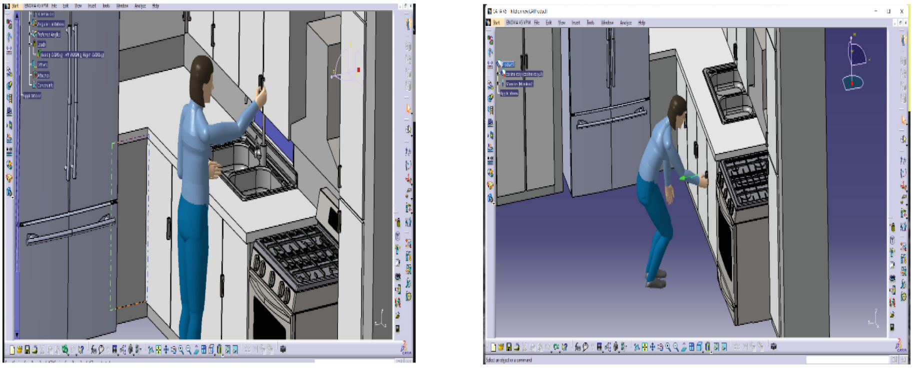


Fig. 2. (**a**) Case 1 opening cupboard. (**b**) Case 2 opening lower cupboard

The reaction on left foot is 285 N and reaction on right foot is 328 N. The load on spine is 1979 N. For turning tap opening as shown in Fig. 3(b), the L4–L5 moment obtained is 14 N m, L4–L5 compression obtained is 2027 N. The total ground reaction obtained is 623 N. The reaction on left foot is 701 N and reaction on right foot is 78 N. The load on spine is 2027 N.

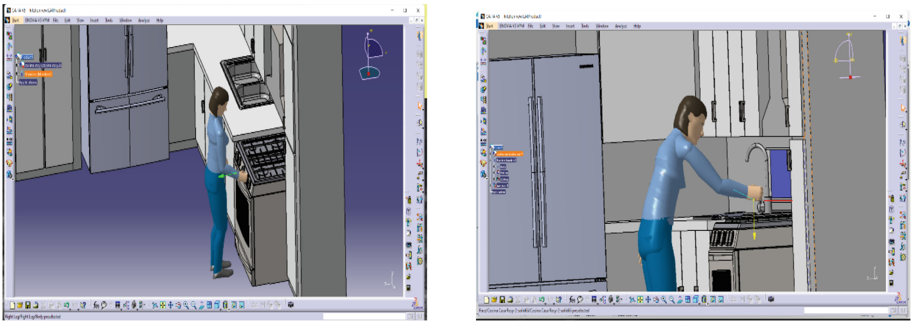


Fig. 3. (**a**) Case 3 gas lighter. (b) Tap opening

For tap closing as shown in Fig. 4(a), the L4–L5 moment obtained is − 20 N m, L4–L5 compression obtained is 1560 N. The total ground reaction obtained is 623 N. The reaction on left foot is 584 N and reaction on right foot is 39 N. The load on spine is 1560 N. For refrigerator door opening as shown in Fig. 4(b), the L4–L5 moment obtained is 33 N m, L4–L5 compression obtained is 2482 N. The total ground reaction obtained is 623 N. The reaction on left foot is 2943 N and reaction on right foot is 2320 N. The load on spine is 2482 N.

For opening bottom drawer as shown in Fig. 5, the L4–L5 moment obtained is 41 N m, L4–L5 compression obtained is 836 N. The total ground reaction obtained is 613 N. The reaction on left foot is 613 N. The load on spine is 836 N. The subsequent analysis is done on new design of kitchen which has certain modification on height and location of refrigerator, tap, gas stove. The ergonomic analysis is conducted and similar parameters

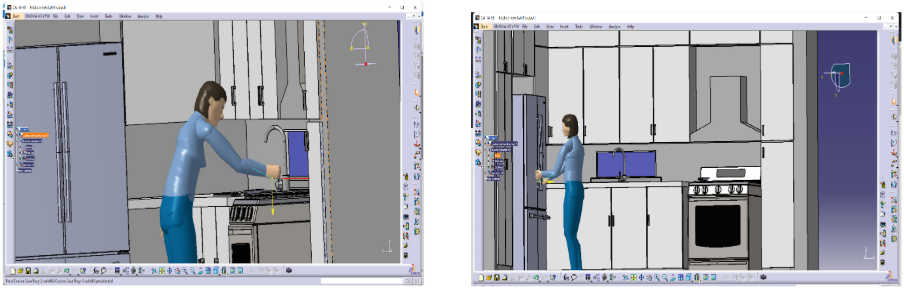


Fig. 4. (**a**) Tap closing. (**b**) Opening refrigerator top door

are evaluated. Figure 6 (a) and (b) shows the analysis of opening lower cupboard and opening middle cupboard, Fig. 7 (a) and (b) shows analysis of turning gas lighter and opening top cupboard shown and Fig. 8 shows analysis of tap opening. Comparison for lower and middle cupboard opening cases & burner operating and overhead cupboard opening cases is shown in Tables 1 and 2 respectively.

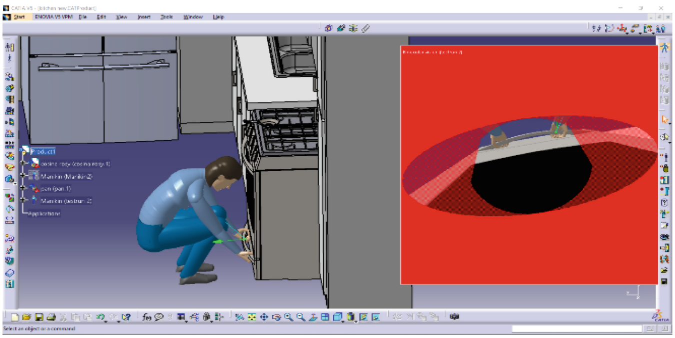


Fig. 5. Opening bottom drawer

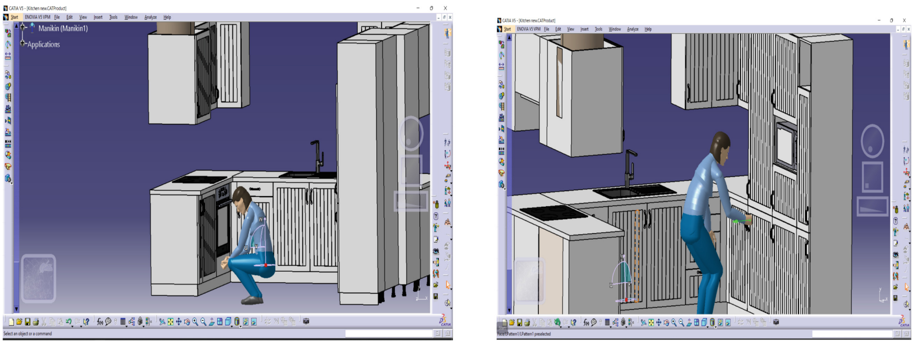


Fig. 6. (**a**) Opening lower cupboard. (**b**) Opening middle cupboard

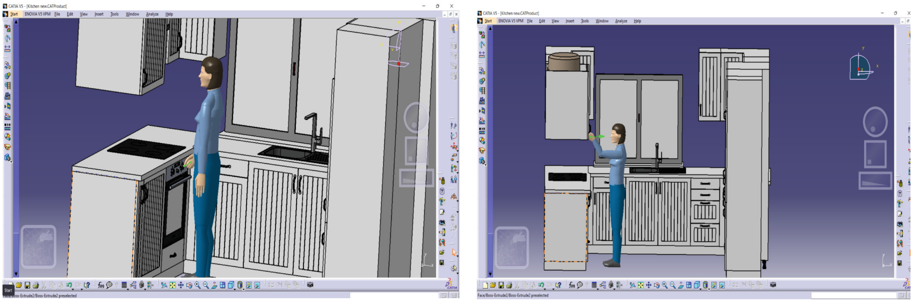


Fig. 7. (**a**) Turning gas lighter. (**b**) Opening top cupboard

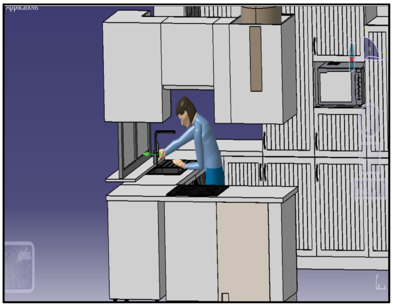


Fig. 8. Tap opening

Earlier studies found that the location of the sink in the kitchen layout had a significant impact on consumers' hygiene practices. Placing sink near to food cooking area would improve hygiene practices [14]. In recent years, the concept of work triangle is used as guideline for planning efficient kitchen [15–18].

The ergonomic design of kitchen can be improved by applying following guidelines.

- Reducing the counter height (platform height) by 24 mm. The new platform height should be maintained to 890 mm in order to reduce pain on L4–L5 joint.
- In earlier design the upper cupboard design was 1320 mm. The upper cupboard should be placed at height 1470 mm to improve ergonomic factor and reduce muscle pain.
- By placing refrigerator on backside of worker the musk skeletal injuries can be mitigated.

Table 1. Lower and middle cupboard opening cases

Lower cupboard opening		
	Case 1	Case 2
L4–L5 compression [N]	836	538
Axial twist compression [N]	50	28
Flex/ext compression [N]	723	99
Spine limit joint shear	130 anterior	114 anterior
Right shoulder flexion-extension	10 extension	4 extension
Middle cupboard opening		
	Old case	New case
L4–L5 compression [N]	1231	2389
Axial twist compression [N]	5	5
Flex/ext compression [N]	1311	231
Spine limit joint shear	126 anterior	
Lumbar (L4–L5) flexion extension	79 extension	7 posterior
Segment proximal force [N] right forearm	3	12

Table 2. Burner operating and overhead cupboard opening cases

Burner operating cases		
	Case 1	Case 2
L4–L5 compression [N]	1979	1973
Axial twist compression [N]	5	5
Flex/ext compression [N]	1875	53
Right shoulder flexion-extension	6 extension	1 extension
Overhead cupboard opening		
L4–L5 compression [N]	2263	875
Axial twist compression [N]	17	16
Flex/ext compression [N]	2144	1027
Spine limit joint shear	136 anterior	116 anterior
Spine limit compression	2263	875
Lumbar (L4–L5) flexion extension	129 extension	62 extension
Segment proximal force [N] right forearm	Z 23	12

5 Conclusion

Using the ergonomic analysis, the kitchen design is analyzed. The analysis results have shown that new kitchen design is ergonomically better than 1st kitchen design.

- For lower cupboard opening, the L4–L5 joint has shown lower compression load as compared to 1st design of kitchen. The percentage reduction of compression is 35.64%.
- For burner opening, the new kitchen design has lower L4–L5 compression as compared to 1st kitchen design. The percentage reduction of compression is 0.3%.
- For overhead cupboard opening, the new kitchen design has lower L4–L5 compression as compared to 1st kitchen design. The reduction of compression load is nearly 3 times.
- For lower cupboard opening, the flex compression is significantly lower as compared to 1st design of kitchen.
- For middle cupboard opening, the flex compression is 6 times lower than 1st design of kitchen.
- For burner opening, the flex compression in new kitchen design is 17 times lower than 1st design of kitchen.
- For overhead cupboard opening, the flex compression of new kitchen design is nearly half of 1st kitchen design.

The future scope of ergonomic design in kitchen is quite promising. As people become increasingly health-conscious and aware of the importance of proper nutrition, the demand for well-designed and functional kitchens will continue to grow. With the advent of the Internet of Things (IoT) and smart home technology, we can expect to see more kitchens that are equipped with sensors, automated appliances, and voice-activated controls. These smart kitchens will be able to adjust lighting, temperature, and humidity levels to optimize cooking conditions, as well as provide personalized recipe suggestions and meal planning. As environmental concerns continue to grow, we can expect to see more sustainable and eco-friendly materials being used in kitchen design. This could include materials like bamboo, recycled glass, and reclaimed wood, which are not only better for the environment but also add a unique aesthetic to the kitchen.

References

1. Patil, A.S., Rajhans, N.R.: Human factors in designing workspace: customizing kitchen counter design. In: Conference Paper in Advances in Intelligent Systems and Computing (2018)
2. Suhas, M., Prajapathi, P., Jolhe, D., Lakhe, R.: A study of Indian kitchen from ergonomic consideration. Ind. Eng. J. (2019)
3. Öztürk, D., Arici, Y.K.: A study on workplace environment ergonomics and conditions of employees in rice factories. Bus. Stud. (2018)
4. Selamat, M.: Ergonomic work system and occupational safety and health performance: mediating effects of psychosocial work factors. Univ. Sains Malays. **62**(1), 27–40 (2016)
5. Shete, K., Pandve, H., Puntambekar, T.: Role of ergonomics in kitchen related back problems. J. Ergon. **5**, 1000–1141 (2015)

6. Middlesworth, M.: Ergonomics 101: the definition, domains, and applications of ergonomics (2020). Retrieved from https://ergo-plus.com/ergonomics-definition-domains-applications
7. Kolus, A., Wells, R., Neumann, P.: Production quality and human factors engineering: a systematic review and theoretical framework. Appl. Ergon. **73**, 55–89 (2018)
8. Lundgren, R.E., McMakin, A.H.: Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks. Wiley (2018)
9. Antonio, G.R.V., Sabanal, R.J.B., Baring, J.A., Sumadlayon, A.R.V.: Ergonomic Practices in University of Cebu–Main Campus (2018)
10. Bhatia, A., Singla, S.: Ergonomic evaluation and customized design of kitchen. Int. J. Innov. Technol. Explor. Eng. **8**(9S), 1033–1039 (2019). https://doi.org/10.35940/ijitee.I1166.0789S19
11. Brito, M.F., Ramos, A.L., Carneiro, P., Gonçalves, M.A.: A continuous improvement assessment tool, considering lean, safety and ergonomics. Int. J. Lean Six Sigma (2019)
12. Singh, D., Thakur, A., Singh, M., Sandhu, A.: IoT implementation strategies amid COVID-19 pandemic. Int. J. Comput. Appl. Technol. **65**(4), 389–398 (2021)
13. Kour, K., et al.: Smart-hydroponic-based framework for saffron cultivation: a precision smart agriculture perspective. Sustainability **14**(3), 1120 (2022). https://doi.org/10.3390/su14031120
14. Mihalache, O.A., Borda, D., Dumitras, L.: Kitchen layouts and consumers' food hygiene practices: ergonomics versus safety. Food Control **131**, 108433 (2022). https://doi.org/10.1016/j.foodcont.2021.108433
15. Adams, C.: What Is the Kitchen Triangle? ThoughCo. (2018). https://www.thoughtco.com/kitchen-work-triangle-1206604
16. Wallender, L.: Use the Kitchen Triangle for Space Planning. The Spruce (2020). https://www.thespruce.com/classic-kitchen-triangle-1822213. Accessed 12 Mar 2021
17. Pehkonen, I., et al.: Evaluation of a participatory ergonomic intervention process in kitchen work. Appl. Ergon. **40**(1), 115–123 (2009). https://doi.org/10.1016/j.apergo.2008.01.006
18. Hagan, J.A., Kwofie, A.S., Baissie, F.: Assessing the effect of kitchen layout on employee's productivity. IJRDO J. Appl. Manag. Sci. **3**(1), 36–45 (2017). ISSN: 2455-9229. https://www.ijrdo.org/index.php/ams/article/view/1053
19. Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022). https://link.springer.com/book/10.1007/978-3-031-19958-5
20. Özalp, B.T.: Determination of domestic kitchen characteristics for elderly Turkish women: a comprehensive ergonomics approach. Open J. Saf. Sci. Technol. **10**, 53–67 (2020)

Author Index

P. Vasant et al. (Eds.): ICO 2023, LNNS 854, pp. 439–441, 2024.
https://doi.org/10.1007/978-3-031-50151-7

Y