ieeeforum INDUSTRIAL ELECTRONICS & ELECTRICAL ENGINEERS FORUM

Proceedings of

ieeeforum

INTERNATIONAL CONFERENCE

02ND MAY 2021 | VENUE: CHENNAI, INDIA

Association With

PROCEEDINGS OF

ieeeforum
INTERNATIONAL CONFERENCE

Date: 02nd May, 2021

Venue: Chennai, India

Organized by

ieeeforum INDUSTRIAL ELECTRONICS & ELECTRICAL ENGINEERS FORUM

Association with

Corporate Address

INSTITUTE OF RESEARCH AND JOURNALS
Plot No. 30, Dharma Vihar, Khandagiri, Bhubaneswar, 751030
Odisha, India
www.iraj.in

Publisher: **Institute for Technology and Research (ITRESEARCH)**

ISBN: 978-93-90150-34-2

Edn: 41

Disclaimer: Authors have ensured sincerely that all the information given in this book is accurate, true, comprehensive, and correct right from the time it has been brought in writing. However, the publishers, the editors, and the authors are not to be held responsible for any kind of omission or error that might appear later on, or for any injury, damage, loss, or financial concerns that might arise as consequences of using the book.

Type set & Printed by:

Institute for Technology and Research (ITRESEARCH)

Khandagiri, Bhubaneswar

About Institute of Research and Journals:

Institute of Research and Journals (IRAJ) is an advanced Non-profit technological forum under Peoples Empowerment Trust, for the Researchers & Scholars "to promote the progress of Science and Technology" by displaying their knowledge in the vicinity of Science and Technology for the service of mankind and the advancement of general welfare.

Objective of IRAJ:

- To provide a world class platform to researchers to share the research findings by organizing International/National Conferences.
- To use the research output of the conference in the class room for the benefits of the students.
- To encourage researchers to identify significant research issues in identified areas, in the field of Science, Engineering, Technology and Management.
- To help dissemination of their work through publications in a journal or in the form of conference proceedings or books.
- To help them in getting feedback on their research work for improving the same and making them more relevant and meaningful, through collective efforts.
- To encourage regional and international communication and collaboration; promote professional interaction and lifelong learning; recognize outstanding contributions of individuals and organizations; encourage scholar researchers to pursue studies and careers in circuit branches and its applications.
- To set up, establish, maintain and manage centers of excellence for the study of /on related subjects and discipline and also to run self supporting projects for the benefit of needy persons, irrespective of their caste, creed or religion.

About ieeeforum:

Industrial Electronics and Electrical Engineers Forum(ieeeforum) is one of the world's largest research forum and Non-profitable professional associations operating under ITR Group meant for research and development in the field of Industrial Electronics and Electrical Engineering. **ieeeforum** is a paramount body that has brought technical revolution and sustainable development in the field of Engineering and technology.

ieeeforum is a forum where innovations & research interest could be supported and developed prioritizing our mutual interest. Our forums & Associates constitutes of Professional leaders, Universities, Organizations & Associations connecting each other with a mission to work as wizards of science for defending the earth.

Conference Committee

Program Chair:

Dr. P. Suresh
Professor, Karpagam College of Engineering,
Coimbatore, Tamil Nadu, India

Prof. R.N Barik
Chairman, IRAJ Research Forum
Mail: chairman@iraj.in

Conference Convener:

Miss. Sumita Nayak
Mob: +91-8280047516

Miss. Gitanjali Pradhan
Mob: +91-9937054932

Publication Head:
Mr. Manas Ranjan Prusty, IRAJ, India

INTERNATIONAL ADVISORY MEMBERS

Prof. Goodarz Ahmadi,
Professor, Mechanical and Aeronautical Engineering,
Clarkson University, USA

Dr Chi Hieu Le,
Senior Lecturer, University of Greenwich.
Kent ME4 4TB. United Kingdom

PROF. (ER.) Anand Nayyar
Department of Computer Applications &
I.T.KCL Institute of Management and Technology,
Jalandhar G.T. Road,Jalandhar-144001, Punjab, India.

Prof. R. M. Khaire,
Professor, Dept. of Elex. and Telecommunication,
B, V University, India

Dr.P.Suresh,
Professor, Karpagam College of Engineering,
Coimbatore, Tamilnadu

Mark Leeson
Associate Professor (Reader)
Area of Expertise: nanoscale communications, evolutionary algorithms, network coding and communication systems

Dr. P. K. Agarwal
Professor, Deptt. of Civil Engineering,
MANIT Bhopal ,Ph. D: IIT Kanpur
M.E: Civil Engg.IIT Roorkee, Membership:
Indian Road Congress (IRC), Institute of Urban Transport (IUT)

Shahriar Shahbazpanahi
Islamic Azad University,
Department of Civil Engineering, Sanandaj,
Kurdistan, Iran, PhD (Structural Engineering),
University Putra Malaysia, Malaysia ,
2009-Present

Harun Bin Sarip
Head of Research and Innovation Dept, UniKL-MICET
Doctorate: Université de La Rochelle, France
Member : International Society of Pharmaceutical Engineer,
Singapore Chapter

Dr.Bilal Ali Yaseen Al-Nassar
The World Islamic Sciences and Education University (WISE)
Faculty of Business and Finance, Department of Management
Information System (MIS), Amman- Jordan

Dr. Md. Al-Amin Bhuiyan
Associate Professor, Dept. of Computer Engineering
King Faisal University, Al Ahssa 31982, Saudi Arabia

Prof. (Er.) Anand nayyar
Department of Computer Applications & I.T.
KCL Institute of Management and Technology,
Jalandhar, G.T. Road, Jalandhar-144001
Punjab, India

Prof. Aleksandr Cariow
Institution or Company: West Pomeranian University of Technology, Szczecin

Dr. P. K. Agarwal
Professor, Deptt. of Civil Engineering,
MANIT Bhopal ,Ph. D: IIT Kanpur
M.E: Civil Engg.IIT Roorkee,
Membership: Indian Road Congress (IRC),
Institute of Urban Transport (IUT)

Dr. VPS Naidu
Principal Scientist & Assoc. Prof., MSDF Lab, FMCD
CSIR - National Aerospace Laboratories,
Bangalore, India

★★★

TABLE OF CONTENTS

★★★

Editorial

Good teaching emanates from Research. The teachers' love for research and their experience in research are vital for the growth of the institution. Any institution is judged by the level and extent of the research work it accomplishes. This sets in a regenerative cycle of excellence. Experience of research leads to quality teaching and quality teaching imparted to the young in turn enriches the research. The campus dynamics needs such type of research teaching research environment.

Technology is the non-linear tool available to humanity, which can affect fundamental changes in the ground rules of economic competitiveness. Science is linked to technology through applications. Technology is linked to economy and environment through manufacture of knowledge products. Economy and environment are linked to technology, which promotes prosperity to the society. We have to use innovation to generate high value added products for becoming a global player. The foundation for academic excellence is the research.

Let us take would like to give, how you young friends can become a great inventors or discoverers. What is the unique nature of thinking minds of discoverers and inventors of the world. "Inventions and discoveries have emanated from creative minds that have been constantly working and imaging the outcome in the mind. With imaging and constant effort, all the forces of the universe work for that inspired mind, thereby leading to inventions or discoveries". Now there are three unique friends to make you great; they are great books, great human beings and great teachers. Teachers should have the capacities to nurture the "creative minds" and "imagining minds".

So this conference has been designed to stimulate the young minds including Research Scholars, Academicians, and Practitioners to contribute their ideas, thoughts and nobility in these disciplines of engineering. It is a pleasure to welcome all the participants, delegates and organizers to this International Conference on behalf of IRAJ Research Forum and ITR family members. This conference has received a great response from all parts of the country and abroad for the presentation and publication in the proceedings. I sincerely thank all the authors for their valuable contribution to this conference. I am indebted towards the Reviewers and Board of Editors for their generous gifts of time, energy and effort for the Conference.

Editor-In Chief

Dr. P. Suresh
Professor, Karpagam College of Engineering,
Coimbatore, Tamil Nadu, India

MEDIATION IN AGRICULTURAL DISPUTES

[1]MEHAK VOHRA, [2]VIJAYLAXMI SHARMA

[1]Research Scholar, Manipal University, Jaipur-Ajmer Express Highway, Dehmi Kalan, Jaipur, Rajasthan 303007, India
[2] Head of Department, Manipal University, Jaipur-Ajmer Express Highway, Dehmi Kalan, Jaipur, Rajasthan 303007, India
E-mail: [1]mehak.vohra1@gmail.com, [2]vijaylaxmi.sharma@jaipur.manipal.edu

Abstract - Agricultural system in India doesn't establish clean contours of an effective judicial system for conflict management. Perhaps, the new farm laws have included certain provisions in response to this need but they seem to harness administrative control and lack judicial touch. Successful attempts at mediation used as dispute resolution in other countries can be referred by the Indian Agriculture System to reduce ever-growing farmer distress.

Keywords - Agricultural Disputes, Agricultural System, Farm Laws, Alternate Dispute Resolution, Mediation, Agricultural Marketing, Judicial System

I. INTRODUCTION

The rising of agriculture after India's independence till date has been commendable. Agriculture saw the light of green and white revolution aiding in overall increased productions and making the country self-sufficient. Despite the progress in primary sector of the economy, the farmers are languishing in poverty and their distress has been soaring. The farmer issues have always been wrapped in political churns and it has never received a well-deserved attention to be resolved. The agricultural legislations are inclusive part of the development process of Indian agriculture and changing the socio-economic conditions of the society. The legislations are structured according to various categories such as land, farm input management, agricultural biotechnology, labour, agricultural marketing, credit & finance, livestock sector and co-operative sector.

Birth of a mightier agricultural system is in process in India with various reforms and agricultural practices budding in one of the most important sectors of the Indian economy. Contract Farming, Zero-Budget Farming, Natural Farming are few of the many trends agriculture is witnessing. Farming is becoming technologically advanced by consolidating various Mandis across the Nation for creating online trading consortium and thereby inviting greater risks. [1]

Most recent one is involvement of private sector entities in the marketing of the agri-produce which was earlier processed through Agricultural Produce Marketing Committees. Though dispute management isn't a new parameter that churns out in this situation but is most definitely is a vital and ambiguous one. Since, the parties involved under such arrangements are farmers on one hand and private entities on the other, it is of utmost relevance to pay attention to the inequality in the status of these parties. Differently fashioned dispute resolution mechanism has to be designed to cater to disparity among the parties and give them enough opportunities of being heard to focus on justice delivery rather than them being caught up in the web of prolonged litigations. The Farm Laws 2020 have further exacerbated on the need to have proper dispute resolution mechanism for farm disputes offering a safe playground for all stakeholders to exercise their associated rights.

Conflict resolution is crucial to prevent injustice, pecuniary loss, violence (in certain cases though not always), barriers to developments in the concerned field at various tiers and also the damage to the worth and efficiency of the judiciary in terms of justice and law enforcement capability. In the light of the legal maxim *lex dilationes semper exhorret* which means 'the law always abhors delays' the major lacuna in the justice delivery system i.e., crisis of delays and backlogs can be tackled with appropriate mediation provisions.

II. DISPUTE RESOLUTION MECHANISM UNDER PRESENT FARM LAWS 2020

Before the enactment of Farmer's Produce Trade and Commerce (Promotion and Facilitation) Act, 2020, all disputes arising out of farm arrangements were addressed under civil court mechanism. Accruing to the difficulty of the farmers to approach court of law for resolution of disputes, the need for a mechanism nearer to farmers was realized. Farmers are the weaker parties in such disputes because of the lack of means to take up the costs of litigation as well as commute to districts with such courts.

At present the Farmer's Produce Trade and Commerce Act, 2020 provides for a mechanism for dispute resolution between farmers and traders. According to this Act an application under section 8(1) has to be made to the Sub-Divisional Magistrate (SDM). The SDM shall further setup a Conciliation Board which shall consist of a chairperson and 2/4 other members. The Chairperson has to be appointed by the SDM and half of the members should represent each of the parties to the dispute. The appointment of these members is made by Chairperson on recommendations of the parties but if parties fail to make the recommendations within 7 days, then the prerogative of appointing the members goes directly to the SDM.

Appellate jurisdiction in these disputes lie with the SDM as per Section 8(5) in case there is no settlement within 30 days. Second appeal lies with the collector or additional collector within 30 days from the order of the SDM. Further, Section 13 of theFarmers (Empowerment and Protection) Agreement on Price Assurance and Farm Services Act 2020 provides for a similar setup for the disputes pertaining to Contract Farming Act. A rather vague system has been developed under Section 9 for issues related to electronic trading and transaction platform. Section 9 provides that the State Government will authorize any officer to take cognizance of the matter and subsequently pass an order within 60 days. The appeal in this circumstance will lie with an officer within 60 days who shall be nominated by the Central Government.

A closer analysis of these provisions in the new farm laws indicate excessive executive control over agri-disputes with absence of judicial involvement. Appointments of the officers as well as the Conciliation Board is done by government which vitiates the core purpose of assisting the parties in an impartial and unbiased manner. The criteria for selection of the Chairman or the qualifications attached to the post have not been mentioned anywhere in the Act which keeps the aggrieved party in a dark space as there is no certainty about the level of knowledge or expertise that the person possesses. A considerable amount of legal acumen is required for the complex contractual and other matters for dispute resolution which evidently is absent in an executive machinery. The decisions given by the conciliators might attract favoritism towards the bigger sharks in the ocean i.e., the private entities or government officials rather than the poor and terrified farmers. In an executive over-reach of power, the Chairman can be appointed as and when the issue arises leaving nobody accountable for the proper functioning of the Board. This results in lack of trust in the authorities dealing with disputes and appointed for resolutions. All of this bars the civil court to have its jurisdiction over any of the matters. According to Swaran Singh Committee Report of1976, disputes relating to revenue, land reforms, ceiling of urban property, procurement and distribution of food grains and other essential commodities shall be decided by a Tribunal.[2]

Since tribunals have existence of both judicial and other technical as well as executive members, a better judicial resolution is possible when farmers want to reach a judicial body. The Committee also recommended that the decisions of the tribunals should be subjected to scrutiny by the Supreme Court under Article 136 of the Constitution of India.[3] On the other hand what appears to be a more suitable solution is to include mediation at the Mandi level or as per new reforms a separate clause for mediation in the agreement itself for a more convenient dispute resolution and settlement for both the parties.

III. BENEFITS OF MEDIATION

The process of mediation is nothing new to the Indian society accruing to the Law Commission Reports and further amendments in the Procedural Codes to recognize and include mediation as a mode of pre-litigation conflict management process. It aims at a better interaction between the disputants by balanced intervention by mediator/mediators. Perhaps, a lot of factors affect the parties to opt for mediation or various approaches available to them. The position of parties, their mindset, complexity of issues involved, expertise and assistance available, financial viability of the disputants, legal awareness are some of the many factors that impact the success of this process. Mediation conceives norms and other specific laws for expected benefits.

Many societies around the world have embraced the use of mediation in dispute resolution and therefore it seems reasonable to consider mediation in solving agri-disputes in India in a more organized way. In Korea, in an attempt to find the theory behind usage of mediation of the disagreements amongst disputants, they believed that mediation is the product of the Confucian harmony-oriented religion. [4] However an equally valid explanation would be that the interdependence of the Korean agrarian society engendered certain norms or in the past there was lack of access to Courts, leading to reliance on outside Court settlements.[5]An investment in a systematic mediation will help in development of new competencies, capacity building and change management.

The debate around mediation being an alternate dispute resolution method is whether the mediators have to be impartial and decide on the merits of the case or they should be partial towards the meeker party to help them protect their rights. The World Bank while making recommendations for practitioners and policy makers has pointed out that mediators should be prepared to deal with contrasting demands of the parties as in some countries along with weak governance there are some added challenges such as corruption and favoritism.[6] Mediation exists in contrary to the formalized and more-professional legal practitioners and their motive to enforce individual rights. Mediation is a mechanism that rides on the theory of solving disputes and when applied in agricultural disputes, it will make sure to upload and maintain the agricultural system.

IV. MEDIATION MECHANISM IN OTHER COUNTRIES FOR AGRICULTURAL DISPUTES

A study of Scandinavian legal and agricultural system points out that the Courts took longer time to come to resolutions as it was more important for them to embed their decisions in people's mind than to come

to a fast resolution. For this the judges would appoint a mediator to investigate problems and to initiate negotiations between the parties.[7]Another series of successful mediation programs can be traced in the U.S Department of Agriculture which prevails in around 20 states in the States to restore failed communication between numerous farmers and lenders for these farmer families to stay in the business. The USDA mediation programs have allowed farm borrowers and their creditors to work out joint solutions, avoiding human suffering and the costly and time-consuming processes of administrative appeal, foreclosure, and litigation.[8]

After the study of mediation programs introduced by the United States Department of Agriculture, one can comprehend that variety of issues can be resolved by the way of mediation. These include conflicts between farmers and their neighbors or communities, credit issues, farms concerns and estate planning, any further disputes affecting the profitability of an agricultural operation. The effectiveness of the mediation program has been calculated by public awareness and availability of mediation program on request and otherwise for farmers and other parties to farm agreements or farm activities. The mediation programs have been linked and supported by the Agricultural Universities, local rural community welfare associations and the Department of Agriculture making it a wholesome program benefiting the farmers not only in resolving disputes at a low cost but also receive other benefits such as better job opportunities, marketing of produce, loan management, reduced economic and social distress, bankruptcy and farmer suicides. By this kind of development of mediation aided farming, the consumer is also largely benefitted and thus making this whole program not only farmer friendly but also consumer friendly.

V. SUITABILITY OF MEDIATION IN INDIAN AGRICULTURE

Under the Farmers (Empowerment and Protection) Agreement on Price Assurance and Farm Services Act 2020, contract farming has been given legal contours in India. This contract farming will involve contracts in the nature of Market Specification Contract, Production Management Contract, Resource Providing Contract etc. For these contracts the parties will be more committed when they are made a part of the drafting process with the help of the mediators instead of imposing the provisions of some standard contracts on them.

The latest Consumer Protection Act 2019 brings out a unique way of involving mediation under the ambit of Redressal Mechanism. The Act concentrates on the establishment of consumer mediation cell attached to the District, State and National Commissions.[9]These cells shall have empaneled mediators and then appropriate mediator shall be nominated amongst the panel considering his suitability for resolving the consumer dispute involved. Likewise, a system can be regulated for agricultural disputes as well by developing synergy between the Agricultural Produce Market Committees (Mandis) and mediators by attaching them to these already existing organizations.

VI. CONCLUSION

The agricultural disputes need to be resolved in a farmer friendly way to reduce their distress. Though the farm laws that have been newly introduced, setup a dispute resolution mechanismbut it has been since then debated vigorously as they undermine the spirit of justice to farmers (the weaker party) in the new-generation farming in India. The implementation of successful mediation mechanism in agri-disputes taking reference from other countries will let us have a sustainable farming culture preventing the farmers from distress. Mediation framework is not supported by a formal legislation in India unlike other forms of alternate dispute resolution systems but the absence of this legislation can be fulfilled by creating specialized mediation process in synergy with the regular working of the different pre-existing agricultural institutions in local areas or the marketing committees in the local limits. The judicial system crisis of backlog of cases and delayed justice can be cured by the use of methods like mediation which are closer to the disputants and help them converse in a more efficient manner.

REFERENCES

[1] https://enam.gov.in/web/eNam-mandi-status Last accessed at 2.54pm on 30/03/2021.
[2] Swaran Singh Committee Report, (1976) 2 SCC (Jour) 45
[3] Ar. 136 of the Constitution of India: Special leave to appeal by the Supreme Court
[4] Notwithstanding anything in this Chapter, the Supreme Court may, in its discretion, grant special leave to appeal from any judgment, decree, determination, sentence or order in any cause or matter passed or made by any court or tribunal in the territory of India
[5] Nothing in clause (1) shall apply to any judgment, determination, sentence or order passed or made by any court or tribunal constituted by or under any law relating to the Armed Forces.
[6] James A. Wall, Jr. John B. Stark and Rhetta L. Standifer, 'Mediation: A Current Review and Theory Development', The Journal of Conflict Resolution, Jun. 2001, Vol. 45, No.3, pg 373.
[7] Ibid.
[8] World Bank, 'Agricultural Innovation Systems – An Investment Sourcebook', 2012.
[9] Jesper Larsson, 'Conflict-resolution mechanisms maintaining an agricultural system', International Journal of the Commons, Vol.10, No.2, 2016, pp 1114
[10] Chester A. Bailey, 'The Role of Mediation in the USDA', Nebraska Law Review, Vol.73, Issue 1, 1994, pp. 142.
[11] Section 74, The Consumer Protection Act, 2019.

★★★

HUMAN ACTIVITY RECOGNITION SYSTEM

[1]PRIYANSHU JAIN, [2]SHRUTI DHANOTIYA, [3]SAMYAK JAIN, [4]SARTHAK RANKA

Information Technology, Medi-Caps University, Indore, India
E-mail: [1]en17it301073@medicaps.ac.in, [2]shruti.dhanotiya@medicaps.ac.in, [3]en17it301083@medicaps.ac.in, [4]en17it301085@medicaps.ac.in

Abstract - Human activity recognition (HAR) has become a popular topic of research because of its wide application. Computers are getting better at solving some very complex problems (like understanding an image) due to the advances in computer vision. Models are being made wherein, if an image is given to the model, it can predict what the image is about, or it can detect whether a particular object is present in the image or not. Here, a deep network architecture using residual bidirectional long short-term memory (LSTM) cells is proposed.Here we uses deep learning for Video Recognition - given a set of labelled videos, train a model so that it can give a label/prediction for a new video.Generally, the proposed network shows improvements on both the temporal (using bidirectional cells) and the spatial (residual connections stacked deeply) dimensions, aiming to enhance the recognition rate.Finally, the confusion matrix of the public domain UCI data set was analyzed.

Keywords - Human Activity Recognition; Bidirectional LSTM; Residual Network; K-Nearest Neighbor, Convolutional Neural Network

I. INTRODUCTION

Human activity recognition (HAR) is of value in both theoretical research and actual practice. It can be used widely, including in health monitoring [1], smart homes [2], and human–computer interactions [3]; for example, LSTM cells are a good choice for solving HAR problems.

The aim of this project is to create a model that can identify the basic human actions like running, walking, standing, sitting, walking_upstairs, walking_downstairs. The model will be given a set of videos where in each video, a person will be performing an action. The label of a video will be the action that is being performed in that particular video. The model will have to learn this relationship, and then it should be able to predict the label of an input (video) that it has never seen. Technically, the model would have to learn to differentiate between various human actions, given some examples of these actions. Unlike traditional algorithms, LSTM can catch relationships in data on the temporal dimension without having to mix the time steps together as a 1D CNN would do.

LSTM architecture can offer great performance and many potential applications. A public domain benchmark of HAR has been introduced, and different methods of recognition have been analyzed [4]. The results showed that the K-Nearest Neighbor (KNN) algorithm outperforms other algorithms in most recognition tasks. Unlike the manual filtering features in previous algorithms, a systematic feature learning method that combines feature extraction with CNN training has also been proposed. Subsequently, DeepConvLSTM networks outperformed previous algorithms in the Opportunity Challenge by an average of 4% of the F1 score; the effects of parameters on the final result were also analyzed. Although researchers have made great strides in HAR, room for improvement remains. Inspired by previous neural networks' architectures, we describe a novel Deep Residual Bidirectional Long Short-term Memory LSTM (Deep-Res-Bidir-LSTM) network. The deep LSTM has improved learning ability and, despite the time required to reach maximum accuracy, shows good accuracy early in training; it is especially suitable for complex, large-scale HAR problems where sensor fusion would be required. Residual connections and bidirectional communication through time are available to ensure the integrity of information flowing deeply through the neural network. In recent years, deep learning has shown applicability to many fields, such as image processing [5], speech recognition [6], and natural language processing [7].

II. RELATED WORK

In ILSVRC 2012, AlexNet [8], proposed by Alex Krizhevsky, took first place, and, since then, deep learning has been considered to be applicable to solving real problems and has done so with impressive accuracy. Indeed, deep learning has become a popular area for scientists and engineers. Another event in 2016 that drew considerable attention was the century man–machine war at the end of the game in which AlphaGo achieved victory. This event also demonstrated that deep learning, based on big data, is a feasible way to solve the non-deterministic polynomial problem. LSTM cells, which were first proposed by JuergenSchmidhuber in 1997 [9], are variants of recurrent neural networks (RNNs). They have special inner gates that allow for consistently better performance than RNN on a time series. Compared with those of other networks, such as CNN, restricted Boltzmann machines (RBM), and auto-encoder (AE), the structure of the LSTM renders it especially good at solving problems involving time

series, such as those related to natural language processing, speech recognition, and weather prediction, because its design enables gradients to flow through time readily. Several experiments were performed with HAR benchmarks: the public domain UCI data set and the Opportunity data set. We compare the accuracy of recognition of our algorithm with those of other algorithm. Finally, we summarize the research and discuss our future work.

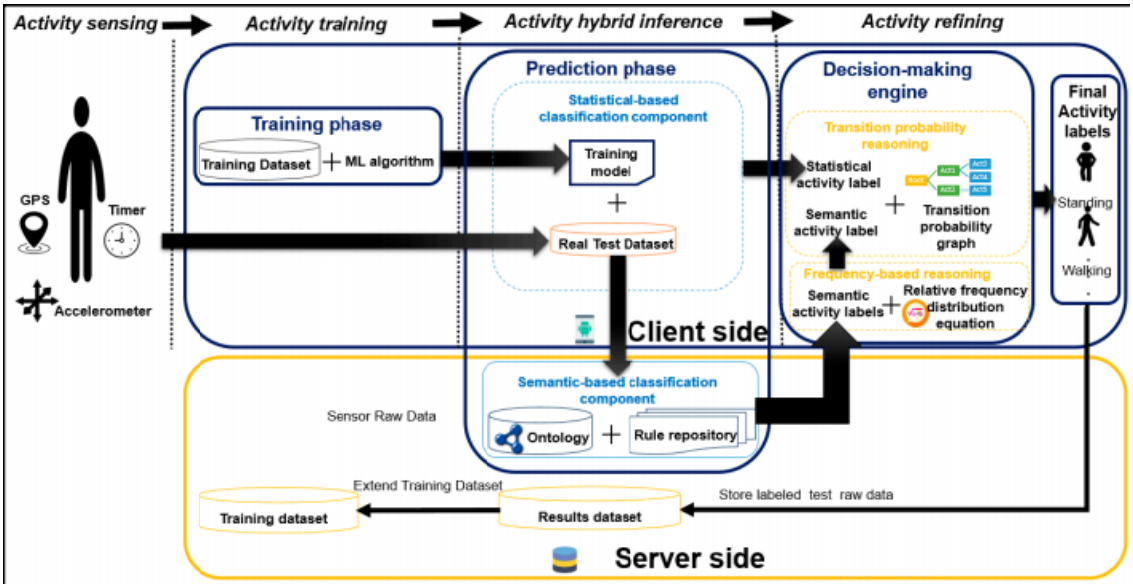


Fig1: Client Server Architecture

III. METHODOLOGY

Data pre-processing

1. Reading in the video frame-by-frame.
2. The videos were captured at a frame rate of 25fps. This means that for each second of the video, there will be 25 frames. We know that within a second, a human body does not perform very significant movement. This implies that most of the frames (per second) in our video will be redundant. Therefore, only a subset of all the frames in a video needs to be extracted. This will also reduce the size of the input data which will in turn help the model train faster and can also prevent over-fitting. Different strategies would be used for frame extraction like:

Extracting a fixed number of frames from the total frames in the video – say only the first 200 frames (i.e., first 8 seconds of the video).

Extracting a fixed number of frames each second from the video – say we need only 5 frames per second from a video whose duration is of 10 seconds. This would return a total of 50 frames from the video. This approach is better in the sense that we are extracting the frames sparsely and uniformly from the entire video.

3. Each frame needs to have the same spatial dimensions (height and width). Hence each frame in a video will have to be resized to the required size.
4. In order to simplify the computations, the frames are converted to grayscale.
5. Normalization – The pixel values ranges from 0 to 255. These values would have to be normalized in order to help our model converge faster and get a better performance. Different normalization techniques can be applied such as:

Min-max Normalization – Get the values of the pixels in a given range (say 0 to 1)

Z-score Normalization – This basically determines the number of standard deviations from the mean a data point is.

Implementation:

One of the most important parts of the project was to load the live video dataset and perform the necessary pre-processing steps. So, we developed a class (Videos) that had a function called (read_videos()) that can be used for reading and processing videos. Creating this was very challenging as we concentrated on generalizing this function for any kind of videos (not specific to this project).We have used NumPy (wherever) for storage and processing of the videos (much faster than in-built python lists with a ton of extra functionalities).The neural network was implemented using Keras.

Experiments

We have tested the LSTM network with the public domain UCI data..Then, we compared it with the outcomes of other methods and analyzed the results. The computer for testing had an i5 CPU with 8 GB RAM as well as an NVIDIA GTX 960m GPU, which has 640 CUDA cores and 8 GB RAM. The GPU and CPU were used alternatively depending on the size of the neural network, which sometimes exceeded the available amount of memory on the graphics card during training.

Data Sets

The research objects of recognition were activities in daily life. Thus, the benchmark for HAR should meet two conditions: first, it should contain most behavioral classes so it reflects real life. Second, it should abstract features and labels for modeling and calculations. Human actions can be divided into several layers such as WALKING, WALKING_UPSTAIRS,WALKING_DOW--NSTAIRS , SITTING , STANDING, JOGGING.A good HAR benchmark should include a clear understanding of the hierarchy.We chose the public domain UCI and the Opportunity data sets for our experiments. The neural network should be readily adaptable to a new data set with an architecture module and a changeable configuration file that also loads the data set. Public domain UCI data set.Each person performed six activities WALKING, WALKING_UPSTAIRS,WALKING_DOWNSTAIRS,SITTING,STANDING, JOGGING The experiments were video-recorded to label the data manually and obtain balanced classes.The dataset obtained was partitioned randomly into two sets: 70% of the dataset were selected for generating the training data, and 30% were selected for generating the test data. Each sample had 561 linear (time-independent) hand-made, preprocessed features from signal analysis (e.g., window's peak frequency), but only six features were used in our study: triaxial gravity acceleration from the accelerometer (from a 0.3 Hz Butterworth low-pass filter) and triaxial body acceleration and triaxial angular velocity from the gyroscope. These are raw signals with a time component and do not fall in the frequency domain but rather in the time domain.

The tasks involved are the following:

Downloading, extracting and pre-processing a video dataset

Dividing the dataset into training and testing data

Create a neural network and train it on the training data.

Test the model on the test data

Compare the performance of the model with some pre-existing models

Metrics: Once the model has been trained on the training data, its performance will be evaluated using the test data. The following metrics will be used:

Accuracy – will be used for evaluating the performance of the model on the test data.

Confusion Matrix - will be used in order to compare the model with the Benchmark model. A confusion matrix is used to describe the performance of a classification model.

Accuracy is the most common evaluation metric used for classification problems. In particular, accuracy is very useful when there are an equal number of samples in each class. Since our dataset have similar characteristics, accuracy would be a suitable metric to evaluate the model.

Benchmark:

The existing models use the notion of local features in space-time to capture and describe local events in a video. The general idea is to describe such events is to define several types of image descriptors over local spatio-temporal neighbourhoods and evaluate these descriptors in the context of recognizing human activities. These points have stable locations in space-time and provide a potential basis for part-based representations of complex motions in video.

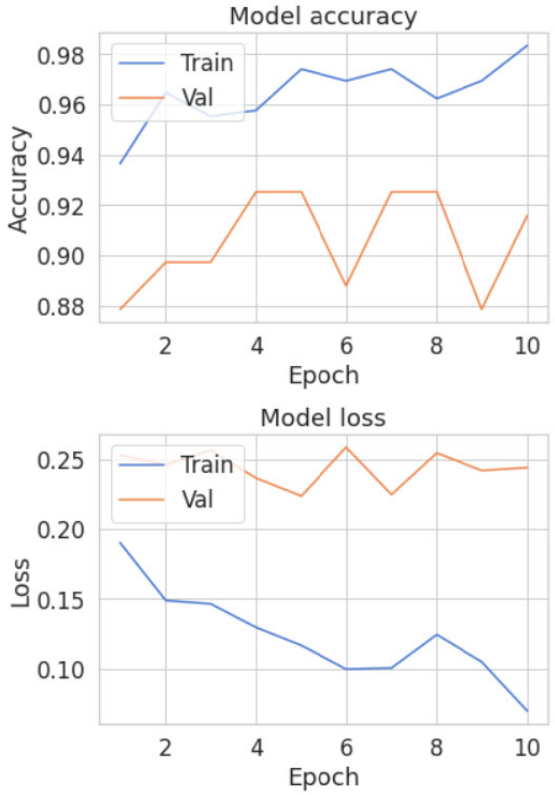


Fig2: Plot curve showing model accuracy and model loss

IV. RESULTS

The six different activities were recognized successfully and the accuracy of model is calculated.

V. CONCLUSIONS & FUTURE SCOPE

In this paper, the significance of HAR research is analyzed, and an overview of emerging methods in the field is provided. LSTM neural networks have been used in many innovations in natural language processing, speech recognition, and weather prediction. This technology was adapted to the HAR task. We proposed the novel framework of the Deep-Res-Bidir-LSTM network. This deep network can enhance learning ability for faster learning in early training. In our experiments, the proposed network was able to improve the accuracy, by 4.78%, for the public domain UCI data set and increase the F1 score, by 3.68%, for the Opportunity data set in comparison with previous work. We also found that window size

was a key parameter. It will be important to find an adaptive way to automatically adjust the searching process and also make the neural network's architecture evolve, such as automatically reshaping, adding, and removing various layers. Also, exploring the effect of mixing 1D time-based convolutions at one or some points in the LSTM cells might improve results. Finally, applying the Deep-Res-Bidir-LSTM network to other fields could be revealing. A good model should have outstanding generalization.

ACKNOWLEDGEMENT

This work was supported by Medicaps University and our guide Ms. ShrutiDhanotiya, Assistant Professor, Medicaps University. We are thankful to our guide and our college for helping us in this journey to reach our ultimate goal.

REFERENCES

[1] Pantelopoulos A, Bourbakis N G. A survey on wearable sensor-based systems for health monitoring and prognosis, IEEE Transactions on Systems Man & Cybernetics Part C Applications & Reviews. 40(1) (2015) 1–12.

[2] Ahmed S H, Kim D. Named data networking-based smart home, ICT Express. 2(3) (2016) 130–134.

[3] Rautaray S, Agrawal A. Vision based hand gesture recognition for human computer interaction: a survey, Artificial Intelligence Review. 43(1) (2017) 1–54.

[4] Chavarriaga R, Sagha H, Calatroni A, et al. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition, Pattern Recognition Letters. 34(15) (2016) 2033–2042.

[5] Dong C, Loy C, He K, et al. Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, 2018.

[6] Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks, in: IEEE International Conference on Acoustics, Speech and Signal Processing. 2017.

[7] Abdel-Hamid O, Mohamed A R, Jiang H, et al. Convolutional Neural Networks for Speech Recognition, IEEE/ACM Transactions on Audio Speech & Language Processing. 22(22) (2014) 1533–1545.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks, in: International Conference on Neural Information Processing Systems, 2012. [9]Hochreiter, Sepp, and Jürgen Schmidhuber. Long short-term memory, Neural computation. 9(8) (1997) 1735–1780.

[9] Anguita, Davide, et al. A Public Domain Dataset for Human Activity Recognition using Smartphones, in: ESANN, 2016. [11]Maclean W J. Spatial Coherence for Visual Motion Analysis, Springer Berlin Heidelberg, 2016.

★★★

DEEP FAKE VIDEO DETECTION USING DEEP LEARNING

[1]NISHITA CHAUHAN, [2]RASHID SAIFI, [3]SHUBHAM CHAUDHARY

[1,2,3]Senior, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042.India
E-mail: [1]nishitachauhan_2k17it77@dtu.ac.in, [2]rashidsaifi_2k17it94@dtu.ac.in, [3]shubhamchaudhary_2k17it117@dtu.ac.in

Abstract - Latest advances and improvements in machine learning algorithms have contributed to the evolution of excellent quality manipulated images which forms video frames, and has a lot of resemblance with the real images which forms video frames. This can have a fatal impact on the way one perceives digitally available knowledge or facts. Current advances in machine learning and camera function have made it possible for images which forms video frames and audio to be manipulated convincingly. These deep-fake videos differ from the real videos via either replacement of audio or the mouth movement or the place where the video was recorded (background). Detection of profound falsehoods with just a little spatial and temporal distortion is particularly difficult. These falsehoods in the distorted videos have now arisen as a threat to humanity and can even serve as war tools in the new digital age. There have been many useful methods and automation procedures for the detection of such deep fake videos. Pose estimation, Facial artifacts, Temporal Pattern Analysis, Background comparison, Eye blinking and Mesoscopic Analysis are some of the techniques used by researchers. We aim to provide a descriptive review of these classical deep Fake detection methods in a distinctive statistical study and hence facilitate the creation and enhancement of a new and a way better approach to handle such deepFake videos through a CNN architecture based on the extreme Inception or Xception model.

Keywords - Deep fake, Neural Networks, Xception

I. INTRODUCTION

The rapid advancement in technology in the fields of computer vision, computer graphics with deep learning has sparked the ability to attain and develop high dimension realistic videos and audios in which a person can be made to do or say anything without his/her own consent. The process thus requires collection of enough sample media files related to the person/object to synthesise a realistic video of the same (Oord et al. 2016) . Surprisingly, there have also been instances where fake videos of people who don't even exist in the living world were developed and spread (Karras et al. 2020)(Karras, Laine, and Aila 2019).

There have been loads of enlightening and entertaining applications via the utilisation of this mechanism but the harms and threat it possesses are even much worse because it can be easily used as a weapon. These false videos and audios created via digital manipulation along with the use of "DeepFake" approaches have turn into a great social concern ("TED: Ideas Worth Spreading" 2008). "DeepFake" refers to the approach which is able to generate hoax videos by interchanging the faces and voices of one individual with another using deep learning. The term was originally used for the first time by a Reddit user who claimed to have evolved an algorithm in which he could swap the faces of real celebrities in other adult videos way back in 2017 ("Deepfakes: What Are They and Why Would I Make One? - BBC Bitesize," n.d.). While faking obscene videos is a concern, there are many other concerns related to these fake videos such as fake news, financial and commercial frauds and hoaxes for social and political reasons. As a result, there have been put forward many researches and growing efforts in the field of general media forensics for detecting such deep fake videos (Swaminathan, Wu, and Liu 2008),(Farid 2009).

Due to the outburst of videos and images posted daily on online social media platforms, it has allowed individuals to create forged videos within minutes together with the aid of deep learning techniques. It is pretty simple to extract huge repositories of videos available on the internet and make real-looking DeepFake media files with deep fake approaches. We can imagine the seriousness of the effect by fake videos, by the fact if something of this sort comes out for a world leader, it could be so disastrous and have even more adverse consequences.

Spotting deep fakes can be related to detecting virus infections, which are a constantly changing problem, it won't be long until the process becomes useless. Hence, we can say that deep fake technology is developing in a manner similar to the virus and antivirus dynamic.

Few years ago, the realism of fake videos was laid back by lack of advacnced tools required for editing, the complex time-consuming process and the domain expertise required. For example, an early work (Bregler et al. 2010) could only forge the motion of the lips of an individual along with an altered audio track and integrating the sound track and the person's face. But lately, many things have changed and advanced in this field of technology. For the current time being we can say, it is very easy for a person to synthesise a fake video to manipulate a real existing face or even a non-existent one.

DEEP-FAKE CREATION

A video that has been altered and manipulated in such a way that it makes a person do or say anything which they never did using deep learning techniques

is called a DeepFake video. The first DeepFake video came to light in 2018 (T. T. Nguyen et al. 2019) and was made using General Adversarial Networks (GAN). The method steered to the creation of specific tools like swap Face2Face, Fake app and Open face which have the capability of generating fake video clips from a large data of images using minimal manipulating operation.

Most of the fake videos involving face swaps are built using general adversarial networks (GAN), including FS- GAN (Nirkin, Keller, and Hassner 2019), Face Swap – GAN ("A Denoising Autoencoder + Adversarial Losses and Attention Mechanisms for Face Swapping.: Shaoanlu/Faceswap-GAN" 2018) and DeepFake FaceSwap ("Faceswap: Deepfakes Software For All," n.d.). The FaceSwap ("MarekKowalski_FaceSwap_ 3D Face Swapping Implemented in Python," n.d.) and Face2Face (Thies et al. 2016) creates fake videos in which an individual's facial expressions and head movements are plotted onto a different individual.

A different unlike approach using text-based synthesis for modifying a video on occurrence of each word(Rossler et al. 2019) rather than requiring a visual or auditory cloning poses even a much bigger challenge for detection of such videos as only a minor change is needed to dramatically alter the significance of the video.

In efforts to those highly manipulated and realistic forged content, many huge steps are being taken out by the research communities to build improved approaches for the detection of fake videos. There are several methods that have been researched upon such as CNN- based, special artifact-based in which individual facial expressions and movements of a specific person using Action Unit (AU) are captured, video forensics, two-stream, auto-encoders, general image manipulation, frequency domain, novel network or module, domain adaptation, metrics learning and using GANs fake-face detection.

II. LITERATURE REVIEW

We will aim to cover the most important associated papers in this study.

In (Ciftci, Demir, and Yin 2019), the authors have talked about various classical methods for face manipulation methods, multimedia forensics. The core contribution of the authors was to provide the FaceForensics++ dataset which extends the preliminary FaceForensics corpus frames until a video ends. Approaches used for generation utilised two state-of-the-art editing methods – Face2Face and FaceSwap (Graphic based approaches), DeepFakes and NeuralTextures (learning based approaches).

In (Li and Lyu 2018), the authors worked on the detection of fake portrait videos using genetic signals. They extracted genetic signals from facial regions on genuine and forged portrait video pairs and applied alterations to calculate the spatial regularity and temporal regularity, captured the signal features in PPG maps and feature sets, and trained a probabilistic SVM and a CNN. They also aggregated genuineness probabilities to decide whether the video is fake or authentic.

In (Yang, Li, and Lyu 2019), the authors worked on to expose deep fake videos by detecting face distorting arti-facts. They argued that algorithms can only produce limited resolution forged images which have to be further processed to match its original image. These kinds of actions leave distinct arti-facts in produced deep fake videos, hence they can be captured by the help of convolutional neural networks (CNNs).

In (H. H. Nguyen, Yamagishi, and Echizen 2019), the study works on uncovering deep fakes through uneven head postures. The process is based on the fact that deep fake videos are generated by breaking combined face regions into the real image, and this process presents errors which can be found when the head postures are characterised from the facial images.

The capsule forensics research in (Do, Na, and Kim 2018) uses capsule networks to identify forged photographs and recordings. A capsule network is used by the framework presented in this paper to detect different forms of fake videos, from repetitive clips via captured videos using deep convolutional neural networks architectures.

In (Rodriguez et al. 2018), the authors used Generative Adversarial Networks (GANs) with Convolutional Neural Networks for forensics face detection. They used GANs to create fake faces with multiple resolutions and sizes to help data augments. Moreover, they applied a deep face recognition system to transfer weight to the system for tough face feature-abstraction.

In (Guera and Delp 2019), in the study of detection of deep fake video manipulation, the authors found that photo response non uniformity (PRNU) analysis is tested for its efficacy at identifying deep fake video alteration. The PRNU analysis showed a noteworthy variation in mean normalised cross correlation scores amongst genuine videos and DeepFakes.

In (Li, Chang, and Lyu 2019), the authors experimented with recurrent neural networks on deep fake video identification. The article suggests a strategy in which deep-fake videos are automatically detected by using a temporal-aware pipeline. To derive frame-level characteristics, their framework uses a convolutional neural network (CNN).

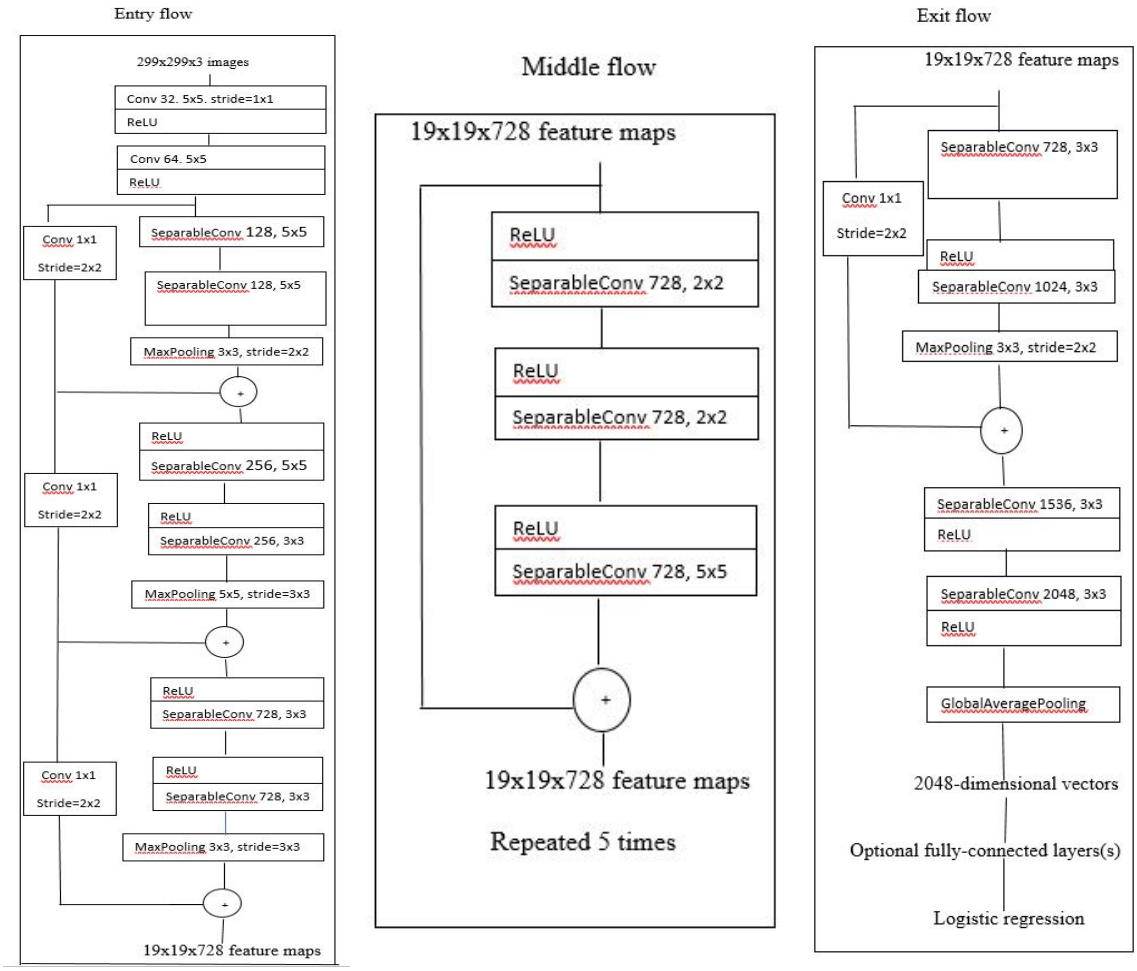
In (Afchar et al. 2019), the approach adopted by the authors was to recognise the absence of human beings' internal physiological signs that are not well captured in the synthesized images. In the synthesis process of fake images, such signals can involve random and automatic physiological processes such as heartbeat, breathing and eye movement, or are sometimes ignored.

In (Sabir et al. 2019), the authors exploited features at a mesoscopic level using 2 networks for fake detection – low number of parameters and target at DeepFakes and Face2Face.
In (Cozzolino et al. 2018), the authors used face orientation as a means for eliminating confounding factors in distinguishing facial changes, instead of using the whole frame and then used bidirectional re-currency rather than just mono-directional. Also, face manipulation generation tools leverage temporal arti-facts and do not impose sequential coherence in the integration process and execute operations on the basis of each frame.
In (Dang et al. 2019), the authors observed that current learning based detecting methods tend to over-fit to specific manipulation that the detectors are trained on, for example detectors trained on Face2Face might fail on FaceSwap or some unseen methods and forensic Transfer handles the problem of identifying classic hidden manipulation approaches, lacking the aid of a large amount of training data.
In (Zhou et al. 2017), the authors worked on fake video recognition based on a CNN, and then accustomed a CNN model for (MANFA) Manipulated (HF-MANFA) Face & A hybrid framework that works on (AdaBoost) Adaptive Boosting and (XGBoost) extreme Gradient Boosting to compact with the unbalanced dataset.

In (Wu, Abdalmageed, and Natarajan 2019), the authors utilised a two-stream network for face alteration detection which comprises of GoogleNet as the triplet network or as a two-stream network architecture and it learns both manipulation arti-facts and local noise residual features.

In (Chollet 2017), the authors unified a deep neural network architecture called the ManTra-Net, which is a simple but efficient self-administered learning job to conceptualise strong image tampering roots from classifying 385 image alteration types which can articulate the tampering localization problem as just a detection of local anomaly problem, then created a Z-score feature to solve the local anomaly, later to assess local anomalies proposed a long short-term memory approach.

III. PROPOSED ARCHITECTURE



In this section, we propose our method, architecture for the detection of deep fake videos. It is a depth-wise separable convolution layers - based CNN architecture. In a CNN architecture, the mapping of spatial and cross-channel correlations is completely decoupled. This theory is a much powerfull version of a Inception architecture, due to which it is named as Xception architecture meaning "Extreme Inception" (Keras Team 2020).

The Inception assumption conveys that a convolution layer tries to acquire filters in a 3D space, containing one channel dimension and two spatial dimensions (height and width), therefore a single convolution kernel performs simultaneous mapping of spatial and cross-channel correlations. So, the Inception model attempts to make the above process easier by wisely dividing it into a number of tasks that would individually work at spatial and cross-channel correlations and hence make it more efficient. Specifically, a classic Inception model first aspects at cross-channel correlations using a set of 1x1 convolutions, maps the input data into 3 or 4 individual spaces that are less significant than the original input space, and then records all correlations in these lesser 3D spaces, using regular 3x3 or 5x5 convolutions. A more simplified version of this Inception module would further formulate a large 1x1 convolution followed by spatial convolutions which would function on non-coinciding sections of the output channels. These observations lead to the fact that it was more sensible to build a much robust hypothesis than the Inception hypothesis, which would have separately mapped spatial correlations and cross-channel correlations.

The extreme version of the Inception model utilises first, a 1x1 convolution to map cross-channel correlations, and then for each output channel, it would distinctly map the spatial correlations. By using the model parameters more efficiently the Xception model marginally out-does Inception V3 on a bigger image classification dataset containing 350 million images and 17,000 classes (Keras Team 2020). The Xception architecture forms the feature extraction base of the network consisting of 36 convolutional layers. These 36 convolutional layers are further arranged into 14 units, the first and last modules of which consist of the linear residual networks around them. This also aids the architecture to be implemented and modified easily, it generally takes very less amount of code if used with a library such as Keras (Keras Team 2020) or TensorFlow-Slim, more like an architecture such as VGG-16 (Amerini et al. 2019).

IV. EXPERIMENTS AND RESULTS

For our experimental analysis we have used 4 sets of data provided by the FaceForensics++ original dataset.

This dataset contains 977 videos taken from youtube, 1000 original extracted sequences which contain a cleared/unhindered face which can be simply traced, it also contains their altered versions via the four methods: Deep Fakes, FaceSwap, Face2Face and NeuralTextures.

FUNCTIONING

By supplying a dlib face to construct a quadratic bounding box, we calculate the bounding box. Then we predict the label of an input image, pre-process the image so that it can be fed into our network, during this process we invoke PIL to cast the image into PIL image. By supplying a dlib face to construct a quadratic bounding box, we calculate the bounding box. Then we predict the label of an input image, pre-process the image so that it can be fed into our network, during this process we invoke PIL to cast the image into PIL image. Outputs are only provided if there is a face, and dlib is used to highlight the face. Our proposed methods use Xception architecture with pre-trained weights from the dataset.

Method Name	Approach	Accuracy
Proposed method	Xception	83%
(Chollet 2017)	XceptionNet Full Image	82.01%
Wang et al. (2019)	GAN-Pipeline Features	80%
(Jafar et al. 2020)	XceptionNet CNN	56.7%

Table 13: The performance of methods

V. CONCLUSION

In this article we have reviewed a lot of methods and approaches towards building a better deep fake video detection model, we have observed various classifiers and frameworks put together for the same reason. After observing the various datasets used by authors and researchers, we chose FaceForensics++ dataset for our analysis with some updation of our own in it.

Most of the authors used inbuilt models, but we developed more improved models of our own using Xception (based on Extreme Inception model) [34] and MesoNet.

In order to train our model on GPU (by using CUDA) the interface is much more easy to use Pytorch instead of tensorflow.

We have achieved that the deep fake detection process works fine with the tuned model of ours according to the needs as compared to the inbuilt one.

The dataset now has every type of deep faked video generated from original video and the dataset is much larger with videos than images.

We worked on videos directly instead of testing and training on images directly.

Our results are based on the probability ratios achieved during the streaming of the video itself, the model keeps on detecting the probability ratio of the fake or real content. The program predicts the content real if the ration of real to fake is greater than equal to 0.5 and else it predicts it to be fake.

FUTURE REFERENCE

In this article, we described a procedure that customs facial mismatches to spot deep-fake videos. With the help of a distinct failure (i.e., triplet), our method first learns characteristics loss which can drive the fake faces away from the actual faces metrically. Then to obtain binary data, we built a Xception classification network. Outcomes show that our approach is operational, and in most situations, it beats most of the classical methods. As there is still viable room for enhancing the efficiency of cross-dataset detection, in the future we would like to propose more powerful face alteration detection techniques which would have a more generalized approach.

REFERENCES

[1] A. van den Oord *et al.*, "WaveNet: A Generative Model for Raw Audio Based on PixelCNN Architecture," *arXiv*, 2016.

[2] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," 2020, doi: 10.1109/CVPR42600.2020.00813.

[3] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," 2019, doi: 10.1109/CVPR.2019.00453.

[4] "TED: Ideas Worth Spreading," *Choice Reviews Online*, vol. 46, no. 04. pp. 46-1805-46–1805, 2008, doi: 10.5860/choice.46-1805.

[5] "Deep Fakes: What are they and why would I make one? - BBC Bitesize." [Online]. Available: https://www.bbc.co.uk/bitesize/articles/zfkwcqt.

[6] A. Swaminathan, M. Wu, and K. J. R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. Inf. Forensics Secur.*, 2008, doi: 10.1109/TIFS.2007.916010.

[7] H. Farid, "Image Forgery Detection A surveyFarid, H. (2009). Image Forgery Detection A survey. IEEE Signal Processing Magazine, 26(2), 16–25. https://doi.org/Doi

[8] 10.1109/Msp.2008.931079," *IEEE Signal Process. Mag.*, 2009.

[9] C. Bregler *et al.*, "Probabilistic Models of Verbal and Body Gestures," in *Computer Vision for Human-Machine Interaction*, 2010.

[10] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep learning for deep fakes creation and detection," *arXiv*. 2019.

[11] Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject agnostic face swapping and reenactment," *arXiv*. 2019.

[12] "A denoising autoencoder + adversarial losses and attention mechanisms for face swapping.:

[13] shaoanlu/faceswap-GAN." 2018, [Online]. Available: https://github.com/shaoanlu/faceswapGAN.

[14] "Faceswap: Deepfakes Software For All." [Online]. Available:

[15] https://github.com/deepfakes/faceswap.

[16] "MarekKowalski_FaceSwap_ 3D face swapping implemented in Python." .

[17] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niebner, "Face2Face: Real-Time Face Capture and Reenactment of RGB Videos," 2016, doi: 10.1109/CVPR.2016.262. [18] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing obama: Learning lip sync from audio," 2017, doi: 10.1145/3072959.3073640.

[18] O. Fried et al., "Text-based Editing of Talking-head Video," ACM Trans. Graph., 2019, doi: 10.1145/3306346.3323028.

[19] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: Learning to detect manipulated facial images," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 1–11, 2019, doi: 10.1109/ICCV.2019.00009.

[20] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals," *arXiv*, vol. X, no. X, pp. 1–17, 2019, doi: 10.1109/tpami.2020.3009287.

[21] Y. Li and S. Lyu, "Exposing DeepFake Videos by Detecting FaceWarping Artifacts," *arXiv*, 2018.

[22] X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2019-May, pp. 8261–8265, 2019, doi: 10.1109/ICASSP.2019.8683164.

[23] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2019-May, pp. 2307–2311, 2019, doi: 10.1109/ICASSP.2019.8682602.

[24] N. Do, I. Na, and S. Kim, "Forensics Face Detection From GANs Using Convolutional Neural Network," no. August, 2018.

[25] A. M. Rodriguez, M. Koopman, A. Macarulla Rodriguez, and Z. Geradts, "Detection of Deep Fake Video Manipulation," *Imvip*, no. December, pp. 133–136, 2018, [Online]. Available: http://www.cost.eu/COST_Actions/ca/CA17124.

[26] D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," *Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, 2019, doi:10.1109/AVSS.2018.8639163.

[27] Y. Li, M. C. Chang, and S. Lyu, "In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking," *arXiv*. 2018.

[28] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," 2019, doi: 10.1109/WIFS.2018.8630761.

[29] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," *arXiv*. 2019.

[30] D. Cozzolino, J. Thies, A. Rössler, C. Riess, M. Nießner, and L. Verdoliva, "ForensicTransfer:

[31] Weakly-supervised Domain Adaptation for Forgery Detection," *arXiv*. 2018.

[32] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, "Face image manipulation detection based on a convolutional neural network," *Expert Syst. Appl.*, 2019, doi: 10.1016/j.eswa.2019.04.005.

[33] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-Stream Neural Networks for Tampered Face Detection," 2017, doi: 10.1109/CVPRW.2017.229.

[34] Y. Wu, W. Abdalmageed, and P. Natarajan, "Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features," 2019, doi:10.1109/CVPR.2019.00977.

[35] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2017, doi:10.1109/CVPR.2017.195.

[36] Keras Team, "Keras - deep learning for humans," *GitHub repository*. 2020, [Online]. Available: https://github.com/keras-team.

[37] "Google AI Blog_ TF-Slim_ A high level library to define complex models in TensorFlow." .

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. "HongguLiu_MesoNet-Pytorch_ The Pytorch Implementation of MesoNet." .

★★★

VOLTAGE MODE FILTER USING VOLTAGE DIFFERENCING TRANSCONDUCTANCE AMPLIFIER

[1]SAKSHAM JAIN, [2]NAMAN JAIN, [3]RAM BHAHAT

[1,2,3]Dept. of Electrical Engineering, Delhi Tech. University, Delhi-110042, India
E-mail: [1]sakshamjain22@protonmail.com, [2]naman2899jain@gmail.com, [3]rambhagat@dtu.ac.in

Abstract - This paper presents a new topology for implementing a tunable single input multi output high pass filter (HPF) and band pass filter (BPF) using the voltage differencing transconductance amplifier (VDTA). The proposed configuration is implemented using single VDTA, two capacitors and one resistor out of which one capacitor is connected to the ground. Simulation has been carried out using PSPICE. The simulated results are compared with the ideal and non-ideal frequency responses of the proposed filter.

Keywords - Voltage Differencing Transconductance Amplifier, Active filter, Analog Signal Processing.

I. INTRODUCTION

The voltage differencing transconductance Amplifier (VDTA)[1] is a multi-output current source controlled by the difference of input voltages which if implemented through CMOS realizations benefit from easy and cheap manufacturing of CMOS circuits[2]. Recently various active filter using current/voltage mode having VDTA as an active component are being designed for analog signal processing [3]-[5].

These active filters are being designed using both Voltage mode [6]-[7] and current mode [8]-[11], out of these two modes the voltage mode filters have an advantage in high frequency operations while the current mode counterparts have an advantage of higher bandwidth. In [6] two VDTAs are used to design all 5 basic filters but suffers from lack of sensitivity analysis while in [7] a universal biquad filter using single VDTA with two capacitors and one resistor is presented. In [8] current mode filter using two capacitors and one resistance is designed while in [9] only two capacitors are used to realize an universal filter configuration, in [10] a single input 3 output biquadratic filter using 2 VDTAs and 2 capacitors is presented. In [11] all three basic filters are realized using 2 capacitors.
In this paper we will be presenting an electronically tunable single input multi output voltage mode filter topology using two capacitors and a single resistor in which one capacitor is grounded which is convenient for physical integration, the presented topology is capable of realizing band pass and high pass configuration at the same time. The working of proposed models has been verified using TSMC CMOS process parameters as defined in Table 1.

II. PROPOSED TOPOLOGY

VDTA is a five-terminal device with symbolic representation as shown in Fig.1. The terminal equations describing a VDTA are given by Eq. (1).

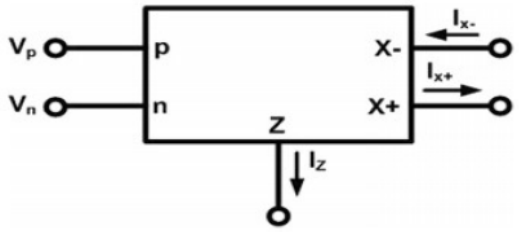


Fig.1. Representation of the VDTA

$$\begin{bmatrix} I_Z \\ I_{X+} \\ I_{X-} \end{bmatrix} = \begin{bmatrix} g_{mF} & -g_{mF} & 0 \\ 0 & 0 & g_{mS} \\ 0 & 0 & -g_{mS} \end{bmatrix} \begin{bmatrix} V_P \\ V_N \\ V_Z \end{bmatrix} \quad (1)$$

The transfer function (TF) of standard second order filters are expressed by Eq. (2).
Here ω_0 shows the cut-off frequency and Q represents the quality factor
Standard filters transfer functions are as given below

$$T(s)_{LP} = \frac{K_0\omega_0}{D(s)}, T(s)_{HP} = \frac{K_0 s^2}{D(s)}, T(s)_{BP} = \frac{K_0 s\frac{\omega_0}{Q}}{D(s)} \quad (2)$$

Where

$$D(s) = s^2 + s\frac{\omega_0}{Q} + \omega_0^2 \quad (3)$$

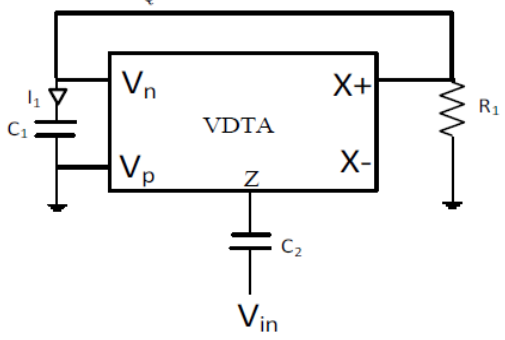


Fig.2.Proposed filter configuration

III. IDEAL ANALYSIS OF PROPOSED FILTER CONFIGURATION

A. BPF REALIZATION

$$\frac{V_N}{V_{in}} = \frac{\frac{g_2}{C_1}s}{s^2 + \frac{s}{C_1R_1} + \frac{g_1g_2}{C_1C_2}} \quad (4)$$

Comparing Eq. (4) with the standard equation for band pass filter we get ω_0 and Bandwidth (BW) as

$$\omega_0 = \sqrt{\frac{g_1 g_2}{C_1 C_2}}\, BW = \frac{1}{C_1 R_1} \quad (5)$$

B. HPF REALIZATION

$$\frac{I_1}{V_{in}} = \frac{g_2 s^2}{s^2 + \frac{s}{C_1 R_1} + \frac{g_1 g_2}{C_1 C_2}} \quad (6)$$

Comparing Eq. (6) with the standard equation for high pass filter we get ω_0 and Q as

$$\omega_0 = \sqrt{\frac{g_1 g_2}{C_1 C_2}}\, Q = R_1 \sqrt{\frac{g_1 g_2 C_1}{C_2}} \quad (7)$$

IV. NON IDEAL ANALYSISOF FILTER CONFIGURATION

The non-ideal analysis of these filters is being carried out by using real model of VDTA where C_{x+} and R_{x+} are parasitic resistance of x+ terminal and C_z and R_z are the parasitic capacitance and resistance of Z terminal and C_n and R_n are the parasitic capacitance and resistance of V_n terminal. β_1 and β_2 are respectively thenon-ideal transfer gain of VDTA. Using straight forward analysis, the non-idealTF and parameters of the filter cut-off frequency ($\overline{\omega_0}$) and non-ideal quality factor ($\bar{Q}$)/Bandwidth ($\overline{BW}$) for these filters are derived and shown in Fig.3.

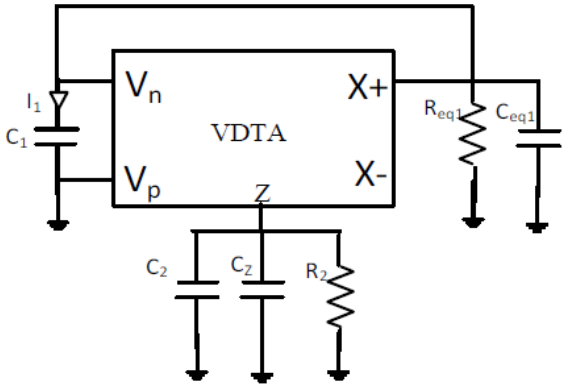


Fig.3.Non-Ideal filter configuration

Here

$$R_{eq1} = R_n \,||\, R_{x+} \,||\, R_1 \quad (8)$$

$$C_{eq1} = C_{x+} + C_n \quad (9)$$

A. Non-Ideal Analysis of BPF

Shown as

$$\frac{V_N}{V_{in}} = \frac{\frac{g_2 \beta_2 C_2}{(C_1 + C_{eq1})(C_2 + C_Z)} s}{s^2 + s\left(\frac{1}{R_{eq1}(C_1 + C_{eq1})} + \frac{1}{R_Z(C_2 + C_Z)}\right) + \frac{1 + g_1 \beta_1 g_2 \beta_2 R_{eq1} R_Z}{R_{eq1} R_Z (C_1 + C_{eq1})(C_2 + C_Z)}} \quad (10)$$

$$\overline{\omega_0} = \sqrt{\frac{1 + g_1 \beta_1 g_2 \beta_2 R_{eq1} R_Z}{R_{eq1} R_Z (C_1 + C_{eq1})(C_2 + C_Z)}} \quad (11)$$

$$\overline{BW} = \left(\frac{1}{R_{eq1}(C_1 + C_{eq1})} + \frac{1}{R_Z(C_2 + C_Z)}\right) \quad (12)$$

B Non-Ideal Analysis of HPF

Shown as

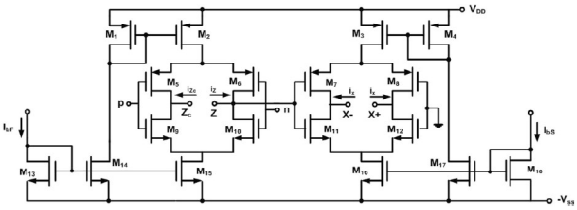


Fig.4. Implementation of VDTA Using CMOS

$$\frac{V_N}{V_{in}} = \frac{\frac{g_2 \beta_2 C_2}{(C_1 + C_{eq1})(C_2 + C_Z)} s}{s^2 + s\left(\frac{1}{R_{eq1}(C_1 + C_{eq1})} + \frac{1}{R_Z(C_2 + C_Z)}\right) + \frac{1 + g_1 \beta_1 g_2 \beta_2 R_{eq1} R_Z}{R_{eq1} R_Z (C_1 + C_{eq1})(C_2 + C_Z)}} \quad (13)$$

$$\overline{\omega_0} = \sqrt{\frac{1 + g_1 \beta_1 g_2 \beta_2 R_{eq}\ R_Z}{R_{eq1} R_Z (C_1 + C_{eq1})(C_2 + C_Z)}} \quad (14)$$

$$\bar{Q} - \sqrt{\frac{(1 + g_1 \beta_1 g_2 \beta_2 R_{eq1} R_Z)(R_{eq1} R_Z (C_1 + C_{eq}\ )(C_2 + C_Z))}{R_{eq1}(C_1 + C_{eq1}) + R_Z(C_2 + C_Z)}} \quad (15)$$

MOS transistors	W/L(μm/μm)
M_1–M_2	15.75/0.25
M_3–M_4, M_9–M_{10}	20.30/0.25
M_5–M_6	14.55/0.25
M_7–M_8	23.30/0.25
M_{11}–M_{12}	5.20/0.25
M_{13}	14.50/0.25
M_{14}	15.50/0.25
M_{15}	18.00/0.25
M_{16}	3.20/0.25
M_{17}–M_{18}	2.80/0.25

Table1. Transistor aspect ratios used in CMOS VDTA circuit

V. SENSTIVITY ANALYSIS

Sensitivity analysis [12] of filters is the study of how the uncertainty in the cut-off frequency, ω_0 and quality factor, Q can be divided and allocated to different sources of uncertainty in its inputs. Sensitivities of ω_0 and Q / BW of these filters for both ideal and non-ideal conditions are derived below.

A. *Ideal Model*

$$S_{C_1}^{\omega_0} = S_{C_2}^{\omega_0} = -\frac{1}{2} \quad (16)$$

$$S_{R_1}^{Q} = 1\, S_{C_1}^{Q} = \frac{1}{2} S_{C_2}^{Q} = -\frac{1}{2} \quad (17)$$

$$S_{C_1}^{BW} = S_{R_1}^{BW} = -1 \quad (18)$$

B. *Non Ideal Model*

$$S_{R_1}^{\overline{\omega_0}} = \frac{-R_{X+} R_1}{R_{X+} R_n + R_1 R_n + R_{X+} R_1 + R_{X+} + R_n + R_{X+} R_n g_1 \beta_1 g_2 \beta_2 R_1 R_Z} \quad (19)$$

$$S_{C_1}^{\overline{\omega_0}} = \frac{-C_1}{2(C_n+C_1+C_{x+})} \quad (20)$$

$$S_{C_2}^{\overline{\omega_0}} = \frac{-C_2}{2(C_2+C_z)} \quad (21)$$

$$S_{C_1}^{\bar{Q}} = \frac{C_1}{2(C_n+C_1+C_{x+})} - \frac{C_1}{2(C_n+C_1+C_{x+}+R_z(C_2+C_z)\left(\frac{1}{R_n}+\frac{1}{R_{x+}}+\frac{1}{R_1}\right)} \quad (22)$$

$$S_{C_2}^{\bar{Q}} = \frac{C_2(C_n+C_1+C_{x+})}{2(C_2+C_z)(C_n+C_1+C_{x+}+R_z(C_2+C_z)(\frac{1}{R_n}+\frac{1}{R_{x+}}+\frac{1}{R_1})} \quad (23)$$

$$S_{R_1}^{\overline{BW}} = \frac{-1}{R_1(\frac{(\frac{1}{R_n}+\frac{1}{R_{x+}}+\frac{1}{R_1})}{(C_n+C_1+C_{x+})+\frac{1}{R_z(C_2+C_z)}})(C_n+C_1+C_{x+})} \quad (24)$$

$$S_{C_1}^{\overline{BW}} = \frac{-C_1(\frac{1}{R_n}+\frac{1}{R_{x+}}+\frac{1}{R_1})}{(\frac{(\frac{1}{R_n}+\frac{1}{R_{x+}}+\frac{1}{R_1})}{(C_n+C_1+C_{x+})+\frac{1}{R_z(C_2+C_z)}})(C_n+C_1+C_{x+})^2} \quad (25)$$

$$S_{C_2}^{\overline{BW}} = \frac{-1}{R_z(\frac{(\frac{1}{R_n}+\frac{1}{R_{x+}}+\frac{1}{R_1})}{(C_n+C_1+C_{x+})+\frac{1}{R_z(C_2+C_z)}})(C_2+C_z)^2} \quad (26)$$

VI. PSPICE SIMULATIONSRESULTS

The working of the proposed filter has been carried out using PSPICE simulation results. For simulating the proposed filter CMOS implementation in Fig4. has been used. The variation of cut-off frequency by varying the circuit parameters verifying the tunability of HPF and BPF is shown.

For BPF, if $C_1 = C_2 = 10$nF, $\omega_0 = 60$ KHz, now when R_1 is varied B.W. also varies while ω_0 stays constant:

R_1	1 KΩ	2KΩ	4KΩ
B.W.	100 KHz	50KHz	25KHz

Whereas for HPF, if $C_1 = C_2 = 10$nF, $\omega_0 = 60$ KHz, now when R_1 is varied Q also varies while ω_0 stays constant:

R_1	1KΩ	1.5KΩ	2KΩ
Q	0.5	0.9	1.2 KHz

The frequency responses of the HPF, BPF have been shown in Fig.5, Fig.6 respectively. From frequency responses we can observe that the simulated frequency responses are in close agreement with the ideal and nonideal frequency responses of these filters.

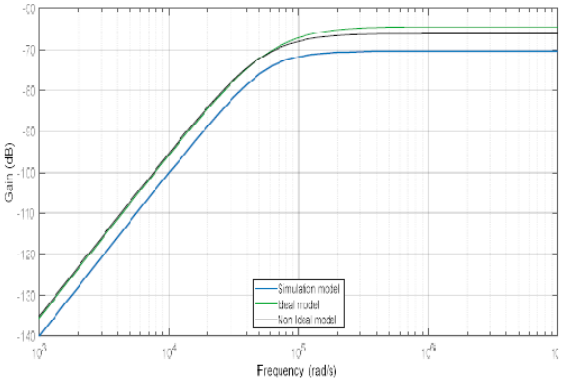


Fig.5.Frequency response of HPF

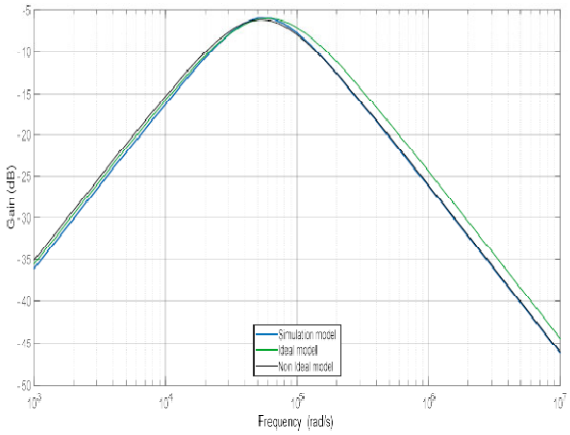


Fig.6. Frequency response of BPF

VII. CONCLUSION

A generalized tuneable filter configuration using single VDTA has been introduced which is capable of realizing HPF and BPF in the voltage mode. The filters have the attractive advantage of having one of the capacitor connected to ground. Moreover the filter circuits can absorb the parasitic effects of VDTA terminals. The simulation results and the frequency response plots have been derived using PSPICE and MATLAB respectively. The simulation results using MATLAB for ideal and non-ideal conditions are in close agreement to each other.

REFERENCES

[1] Biolek, R. Senani, and V. Biolková, "Active Elements for Analog Signal Processing: Classification, Review, and New Proposals," vol. 17, no. 4, p. 18, 2008.

[2] N. H. E. Weste and K. Eshraghian, "Principles of CMOS VLSI design: A systems perspective," NASA STI/Recon Technical Report A, vol. 85, 1985, Accessed: Apr. 15, 2021.

[3] Ye, "New Simple CMOS Realization of Voltage Differencing Transconductance Amplifier and Its RF Filter Application," vol. 20, no. 3, p. 6, 2011.

[4] Prasad and D. R. Bhaskar, "Electronically Controllable Explicit Current Output Sinusoidal Oscillator Employing Single VDTA," ISRN Electronics, vol. 2012, pp. 1–5, Sep. 2012, doi: 10.5402/2012/382560.

[5] Satansup, T. Pukkalanun, and W. Tangsrirat, "Electronically Tunable Single-Input Five-Output Voltage-Mode Universal Filter Using VDTAs and Grounded Passive Elements," Circuits Syst Signal Process, vol. 32, no. 3, pp. 945–957, Jun.2013, doi: 10.1007/s00034-012-9492-0.

[6] G. Gupta, S. V. Singh, and S. V. Bhooshan, "VDTA Based Electronically Tunable Voltage-Mode and Trans-Admittance Biquad Filter," Circuits and Systems, vol. 06, no. 03, Art. no.03, Mar. 2015, doi: 10.4236/cs.2015.63010.

[7] Prasad, D. R. Bhaskar, and M. Srivastava, "Universal voltagemode biquad filter using voltage differencing transconductance amplifier," IJPAP Vol.51(12) [December 2013], Dec.2013, Accessed: Apr. 15, 2021.

[8] Prasad, D. R. Bhaskar, and M. Srivastava, "Universal Current- Mode Biquad Filter Using a VDTA," CS, vol. 04, no. 01, pp. 29–33, 2013, doi: 10.4236/cs.2013.41006.

[9] Satansup and W. Tangsrirat, "Compact VDTA-based currentmode electronically tunable universal filters using grounded capacitors," Microelectronics Journal, vol. 45, no. 6, pp. 613– 618, Jun. 2014, doi: 10.1016/j.mejo.2014.04.008.

[10] Prasad, M. Srivastava, and D. R. Bhaskar, "Transadmittance Type Universal Current-Mode Biquad Filter Using VDTAs," International Scholarly Research Notices, vol. 2014, pp. 1–4, Aug. 2014, doi: 10.1155/2014/762845.

[11] J. Satansup and W. Tangsrirat, Single VDTA-Based Current-Mode Electronically Tunable Multifunction Filter. 2012

[12] Behzad Razavi, ", Fundamentals of Microelectronics", 2nd edition, Wiley, 2006, pp. 673-677.

★★★

BLOCKCHAIN-BASED PUBLIC UPRIGHTNESS CHECK FOR DISTRIBUTED STORAGE AGAINST HESITATING INSPECTORS

[1]HARITHAA S, [2]ANBARAZSI T, [3]LAVANYA T, [4]JACKULIN C

[1,2,3]Computer Science and Engineering, Panimalar Engineering College, Anna University, Chennai, India
[4]Assistant Professor, Department of CSE, Panimalar Engineering College, Anna University, Chennai, India
E-mail: [1]harithaa2708@gmail.com, [2]tanbarazsi18@gmail.com, [3]lavanyathangamani12@gmail.com,
[4]chin.jackulin@gmail.com

Abstract - The network storage services achieve widespread adoption, security and performance issues are getting primary concerns, affecting the scalability of storage systems. Most of the cloud users purchase cloud storage "as pay as you go" consistent with their data volume. we propose a blockchain-based deduplicatable data auditing mechanism, to scale back the wastage of cloud storage resources and to perform uprightness checks for data. The prevailing system with a deduplication mechanism cannot solve the issues of security and storage, it requires high cost and therefore the reliance on trusted third parties. Hence our proposed system consists of a client-side data deduplication scheme supported a bilinear pairing cryptosystem to scale back the value and increase the safety for users and repair providers. The cloud service provider (CSP) doesn't allow the owner to upload the duplicate files but later to the authentication of the owner, CSP can allow the owner to form the updating of files that are uploaded within the cloud to avoid storage waste. The info outsourcing, and auditing processes also verified by the service provider to enhance the safety issues, and to watch the unreliable third-parties, a blockchain system is employed to record those behaviours of both entities

Keywords - Network Storage, Deduplication, Data Auditing, Blockchain

I. INTRODUCTION

In network computing era, network storage services achieve widespread adoption and benefit countless users worldwide thanks to the superior capability of providing low cost and highly scalable storage anytime, anywhere. consistent with a recent report, by 2025, approximately 50% of information or data are going to be stored in network storage devices with a fantastic market price of quite $100 billion globally.

This is thanks to the very fact that network storage service providers may maliciously or accidentally corrupt the user's data. Remote data auditing mechanism is one among the only but best security solutions which will help to see the integrity of outsourced data has attracted much attention. Meanwhile, because almost 75% of outsourced data are duplicate copies, the deduplication technique that eliminates duplicate data is widely adopted within the commercial settings for reducing storage costs and for improving the scalability of the system.Numerous works of data auditing and deduplication are proposed, for the sake of ensuring the integrity of outsourced data and improving the efficiency and scalability of network storage services.

A public key encryption (PKE) scheme such that when being implemented in a bilinear group pairing techniques, anyone is able to check whether two ciphertexts are encryptions of the same message is proposed. Bilinear map operations are not required in key generation, encryption or decryption procedures of the PKE scheme, but is only required when people want to do an equality test (on the encrypted messages) between two cipher texts that may be generated using different public keys. It shows that our PKE scheme can be used in different applications such as searchable encryption and partitioning encrypted data. Moreover, implementing in a non-bilinear group, the security of our PKE scheme can be strengthened from One-Way CCA to a weak form of IND-CCA.

However, Distributed computing empowers clients to eliminate the need of the need of neighborhood equipment design, which eliminates the weight of the clients from high calculation costs. Hence, to ensure clients' protection, information is typically encrypted before being shipped off the cloud worker and the framework is unusable. The confidential data in an enterprise may be illegally accessed through a remote interface provided by a multicloud, or relevant data and archives may be lost or tampered with when they are stored into an uncertain storage pool outside the enterprise. Therefore, it is indispensable for cloud service providers (CSPs) to provide security techniques for managing their storage services.[1] .

As a crucial classic scheme, the Provable Data Possession (PDP) protocol is taken into account effective in protecting the integrity of outsourced data and has inspired many subsequent data auditing schemes. within the classic PDP protocol, the info of the user are going to be pre-processed to get verification metadata before uploaded to the service provider, which enables users can remotely check data integrity without the dependency on local copies . Nevertheless, these schemes fail to integrate the characteristics of enormous duplicate data in cyberspace fully, and there's space for improving their efficiency

further. Integrity verification and protection is an active research area; numerous research problems belong to this area have been studied intensively in the past.[2]

As an efficient thanks to improve the scalability and efficiency of network storage services, deduplication technology enables a service provider to save lots of just one copy of multiple same data, thereby significantly reducing his storage overhead. Data owners who lack mutual trust will independently generate verification metadata for the duplicate file, which imposes unnecessary overhead, when data auditing is trivially performed during a network storage service that performs data deduplication To unravel this problem, some de-duplicatable data auditing schemes use validation metadata which will be shared across multiple tenants. However, most of those approaches believe a trusted Third-Party Auditor to help users in auditing, which can be difficult to seek out within the network storage services. Additionally , multiple users may repeatedly check the integrity of knowledge that they jointly hold, which can also cause unnecessary overhead. The benefits of deduplication include reduced infrastructure costs, reduced management costs, many cloud storage providers such as Dropbox, Memopal and Mozy use client side deduplication in order to save resources which results in avoiding storage of redundant data in cloud storage servers and network bandwidth savings by eliminating transmission of same contents several times.[3]

The recent rise of blockchain technology has promoted the event of the many industries, and also offers new techniques.Through the chain arrangement and distributed storage, the blockchain provides a distributed tamper resistant to help users interact, and makes it possible for credible cooperation between strange entities However, the combination of the above methods remains faraway from perfect and brings unnecessary extra overhead to the server. An efficient and reliable network storage service model designed from a multitenant perspective remains desired. To ovwrecome security issues we propose blockchain-enabled de-duplicatable data auditing mechanism

II. RELATED WORK

There are many data auditing schemes have been proposed to guarantee the integrity of users outsourced data in network storage services.Sanjay Kumar Madria, designed Security and privacy of big data in a cloud environment in which he points, in smart meters, data of the consumers must be protected else private information are often leaked. Similarly, the cost-efficiency, reduced overhead management and dynamic resource needs, content owners are outsourcing their data to the cloud who can act as a service provider on their behalf[4].However, by outsourcing their data to the cloud, the owners may lose access control and privacy of knowledge as cloud becomes a third party. By using these data storage services, owners can relieve the burden of local data storage and maintenance.

Since data owners and therefore the cloud servers aren't within the same trusted domain, the outsourced data could also be in danger because the cloud server may not be fully trusted. Therefore, data integrity is of critical importance. Cloud should let the owners or a trusted third party to see for the integrity of their data storage without demanding an area copy of the information .Dharani P, Berlin in their paper on Survey on secret sharing scheme with deduplication in cloud computing says that data deduplication is one among the techniques used for eliminating duplicate copies of data which is widely utilized in cloud to scale back space for storing and increase bandwidth. Convergent encryption has been extensively adopted for secure deduplication, so as to use efficiently and reliably manage an enormous number of convergent keys. A baseline approach named as Dekey is to distribute the convergent key which might be shared across multiple servers.

A heavy computational cost is required to form n shares and recover the key as an answer to the present problem. Hence a replacement (k, L, n)-threshold ramp scheme is proposed which is ideal , idle and faster secret sharing scheme, every combination of k or more participants can recover the key , but every group of but k participants cannot obtain any information about the key[5].Jing Chen in this model named DeyPoS(Deduplicatable Dynamic Proof of Storage) for Multi-User Environments states dynamic Proof of Storage (PoS) may be a useful cryptographic primitive that permits a user to see the integrity of outsourced files and to efficiently update the files during a cloud server. we introduce the concept of deduplicatable dynamic proof of storage and propose an efficient construction called DeyPoS, to realize dynamic PoS and secure cross-user deduplication. Considering the challenges of structure diversity and personal tag generation, we exploit a completely unique tool called Homomorphic Authenticated Tree (HAT).

However, A practical multi-user cloud storage system needs the secure client-side cross-user deduplication technique, which allows a user to skip the uploading process and acquire the ownership of the files immediately, when other owners of an equivalent files have uploaded them to the cloud server[6].

The rise of blockchain technology has also promoted the development of data deduplication and data uprightness check mechanism. In Cooperative and Distributed Computation Offloading for Blockchain-Empowered Industrial Internet of Things model scheme Wuhui Chen, Zhen Zhang, Zicong Hong realizes, amultihop cooperative and distributed computation offloading algorithm that considers the data processing tasks and the mining tasks together for blockchain-empowered IIoT[7].Y. Yu, Y. Li, B. Yang, W. Susilo, G. Yang, and J. Bai in Attribute-based cloud data integrity auditing for secure outsourced storage, proposed users can upload files to cloud through some customized attribute set and specify some designated auditor set to see the integrity of the outsourced data. The system formalize the system model and therefore the security model for this new primitive, and describe a concrete construction of attribute-based cloud data integrity auditing protocol[8]. The new protocol offers desirable properties namely attribute privacy-preserving and collusion-resistance. It prove soundness of protocol supported the computational Diffie-Hellman assumption and therefore the discrete logarithm assumption. Finally, develop a prototype of the protocol which demonstrates the practicality of the protocol.

Data integrity, one among the foremost burning challenges in secure cloud storage. Outsourced data auditing protocols enable a verifier to efficiently check the integrity of the outsourced files without downloading the whole file from the cloud, which may dramatically reduce the communication overhead between the cloud server and therefore the verifier. Existing protocols are mostly supported public key infrastructure or a particular identity, which lacks flexibility of key management.P. Puzio, R. Molva, M. Onen, and S. Loureiro in deduplication with encrypted data for cloud storage[9] says,a secure and efficient storage service which assures block-level deduplication and data confidentiality at an equivalent time.

Although supported convergent encryption, ClouDedup remains secure because of the definition of a component that implements a further encryption operation and an access control mechanism. Furthermore, because the requirement for deduplication at block-level raises a problem with reference to key managementto incorporate a replacement component so as to implement the key management for every block alongside the particular deduplication operation. But, any potentially untrusted cloud storage provider like Amazon, Dropbox and Google Drive, can play the role of storage provider. As a part of future work, ClouDedup could also be extended with more security measures like proofs of retrievability , data integrity checking and search over encrypted data.J. Shen, D. Liu, M. Z. A. Bhuiyan, J. Shen, X. Sun, and A. Castiglione in their Secure verifiable database supporting efficient dynamic operations in cloud computing implemented a secure verifiable database scheme that's supported the polynomial commitment for cloud computing, which may realize the verifiability of database records within the cloud. Moreover, the proposed scheme can support public verifiability therein all clients within the system can verify the database. Additionally , It use the BLS signature and therefore the index-hash table to construct dynamic operations for the database. Security analysis shows that scheme are able to do real-world security requirements. The simulation results show that scheme is more efficient than similar schemes. the safety analysis shows that scheme are able to do the properties of security, correctness, verifiability and accountability. within the performance analysis, It compare the scheme with two similar schemes. The comparison results and simulation results indicate that scheme is more efficient than similar schemes, which demonstrates that scheme are often well utilized in secure verifiability for databases in cloud computing[10].

The model for provable data possession at untrusted stores, designed a challenge response by G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song.In that,when the data is stored within the cloud memory device , an extended time, enterprises and users inevitably will have security concerns, fearing that the knowledge is really stored within the cloud remains within the memory device or too long without access to, has long been the cloud server removed or destroyed, leading to businesses and users within the future can’t access or restore the info files. Therefore, this scheme goal to research and style for data storage cloud computing environments that are proved. Stored within the cloud for data storage, research and develop a security and efficient storage of proof protocol, can also delegate or authorize others to public verifiability whether the info actually stored within the cloud storage devices[11].

It can provable data possession within the model, which reduce the info block access, but also reduce the quantity of computation on the server and client and server traffic.. It exceeds what did within the past, the development has delivered to the bandwidth, computation and storage system. And it applied the general public (third party) verification.
In Secure and constant cost public cloud storage auditing with deduplication[12],J. Yuan and S. Yu introduce a novel scheme based on techniques including polynomial-based authentication tags and homomorphic linear authenticators. The design allows deduplication of both files and their corresponding authentication tags. Data integrity

auditing and storage deduplication are achieved simultaneously. Proposed scheme is also characterized by constant realtime communication and computational cost on the user side. Public auditing and batch auditing are both supported. Hence, proposed scheme outperforms existing POR and PDP schemes while providing the additional functionality of deduplication. It proves the security on Computational Diffie-Hellman problem, the Static Diffie-Hellman problem and the t-Strong Diffie-Hellman problem.

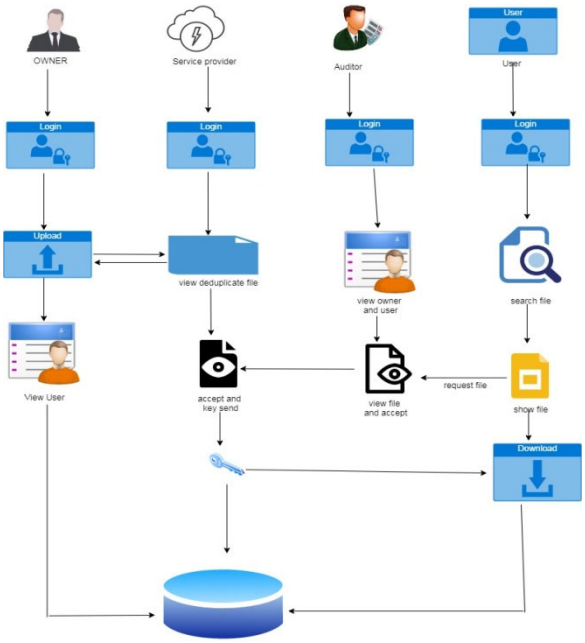
The Secure and efficient proof of storage with deduplication proposed by Q. Zheng and S. Xu[13], introduces a replacement primitive called μR-MLE2 which provides a partial positive account this challenging problem. It propose two schemes: static scheme and dynamic scheme, where the latter one allows tree adjustment by increasing some computation cost. Its main trick is to use the interactive protocol supported static or dynamic decision trees. The advantage gained from it's , by interacting with clients, the server will reduce the time complexity of deduplication equality test from linear time to efficient logarithmic time over the entire data items within the database. the safety analysis and performance evaluation show that schemes are Path-PRV-CDA2 secure and achieve several orders of magnitude higher performance for data equality test than R-MLE2 scheme when the amount of knowledge items are relatively large.

To achieve secure and efficient data deduplication, construct two interactive schemes supported static and dynamic deduplication decision tree structures, respectively. The static deduplication decision tree is made supported the random elements from the client, which doesn't allow the tree to update. However, the dynamic deduplication decision tree is made supported the designed self-generation tree, which allows the server to conduct tree update and a few other optimization.

III. PROPOSED WORK

Inspired by the above challenges, this paper proposes a blockchain-enabled de-duplicatable data auditing mechanism to meet the security requirements while reducing the extra overhead caused by data duplication, thereby improving the scalability of the system. In our scheme, a client-side deduplication mechanism based on bilinear maps is used to reduce the transmission and storage overhead of duplicate data in the data uploading process.

For duplicate files, the user only needs to provide proof of ownership to the service provider without repeatedly uploading them or to re-upload them. To check the integrity of users ,The TPA will respond to the verification requests from users and requires the service provider to return an integrity proof with the challenge and specified data blocks. TPA will verify the proof and make the audit results public on the blockchain. To download the file ,the users need to enter the file key and CSP key provided by the service providers. Blockchain will record information of data outsourcing and data auditing to guide TPA to avoid double-checking of files and form audit logs for user supervision to ensure the correctness of audit results



IV.MODELS AND GOALS

System Model

The system model of our scheme includes five sorts of entities: the user, the network storage service provider (SP), the TPA, the blockchain nodes, and therefore the system manager. they're defined as follows:

- The user uses the network storage service to outsource his data to the SP, and he also hires a TPA to see the integrity of the outsourced data. Different users don't trust one another and should outsource an equivalent data. A malicious user may attempt to spoof the SP to urge data from other users.

- The SP provides network storage services to assist users store their data. SP also uses the deduplication technique to get rid of duplicate data. However, SP could also be untrustworthy and corrupt user data.

- The TPA has significant computing resources and can help users audit data integrity, but the TPA isn't completely trusted.

- The blockchain node may be a node within the blockchain network. All nodes are liable for maintaining blockchain data and handling blockchain transactions.

- The system manager is an entity liable for generating some parameters.

Definition 1. (blockchain-enabled de-duplicatable data auditing mechanism): The blockchain-enabled de-duplicatable data auditing mechanism consists of three algorithms (Setup, Upload, Check Log) and two interactive protocols (De-dup, Audit). In order of the workflow, they can be defined as follows:

(1) Setup(k) → (*sk, Para*). Based on the security parameter k, the Setup algorithm generate user's secret key *sk* and a public set of system parameter *Para.*

(2) Upload(*Para, F, sk*) → γ. From the input public parameter set *Para,* user's file *F* and private key *sk*, the algorithm computes a tag set γ as the verification metadata.

(3) Dedup. The interactive protocol achieves proof of ownership through the following three algorithms.

- DedupChallen(*Para*) → *(*$chal_s$). Based on the public parameter set *Para*, the algorithm generates a deduplication challenge nonce $chal_s$.

- DedupProofGen(*Para,* $chal_s$*, F*) → ($proof_s$). From the input public parameter set Para, a challenge nonce $Chal_s$ and the file *F*, the algorithm generates a proof of ownership $proof_s$.

- DedupVerify(*Para,* $proof_s$) → 0/1. Based on the public parameter set *Para* and the proof of ownership $proof_s$, the algorithm output the verification result which is 0 for rejection and 1 for acceptance.

(4) Audit. The interactive protocol achieves proof of data possession through the following three algorithms.

- AuditChallen(*Para*) → ($chal_t$). Based on the public parameter set *Para* , the algorithm generates an integrity challenge nonce $Chal_t$.

- AuditProofGen(*Para, chalt, F*) → ($proof_t$). From the input public parameter set *Para*, a challenge nonce *chalt* and a file F, the algorithm generates an integrity proof $proof_t$.

- AuditVerify(*Para*, $proof_t$) → (0/1, Log). Input the public parameter set *Para* and the integrity proof $proof_t$, the algorithm outputs the audit result in terms of 0 or 1, and the audit Log *Log*.

(5) CheckLog(*Log*) → 0/1. According to public parameter set *Para* and the TPA's audit log *Log*, this algorithm outputs 0 or 1 to indicate whether the audit result is correct.

Threat Model

In this research, we assume that the majority of the entity participants including service providers, users, and TPAs, are untrustworthy. An untrusted service provider may drop the user's outsourced data for cost-saving and check out to trick the TPA and therefore the user through integrity verification to concealing the loss or corruption of stored data. A user may misuse the client-side deduplication mechanism to spoof to get outsourced data stored by the service provider fraudulently. Similarly, a TPA may collude with the server and not perform integrity verification consistent with the user's requirements.

In addition, users with an equivalent file won't blindly trust each other's audit results of the file and can conduct repeated audits. However, all involved participants are assumed rational, i.e., they're going to not cheat or collude even when their own interests are violated. additionally, the blockchain technique is taken into account reliable and most of the nodes within the blockchain network are trusted, which may be guaranteed in most practices. Therefore, the records on the blockchain can't be tampered with, which eliminates the likelihood of malicious modification of knowledge by participants within the model. Besides, we assume that the adversary cannot crack the involved cryptographic technique within the scheme directly. We also assume that every entity can communicate securely off-chain, and therefore the files and keys passed won't be

V. MODULES

- USER INTERFACE DESIGN
- FILE UPLOAD
- SERVICE PROVIDER LOGIN
- REQUEST ACCEPT AND RE UPLOAD
- USER REQUEST
- RESPONSE FROM THIRD AND SP SEND KEY
- DOWNLOAD FILE

User Interface Design:

This is the primary module of our project. The important role for the user is to maneuver login window to user window. This module has created for the security purpose. During this login page we need to enter login user id and password. it'll check username and password is match or not (valid user id and valid password). If we enter any invalid username or password, we can't enter into login

window to user window it'll shows error message. So, we are preventing from unauthorized user getting into the login window to user window. it'll provide an honest security for our project. So, server contain user id and password server also check the authentication of the user. It well improves the safety and preventing from unauthorized user enters into the network. In our project we are using JSP for creating design. Here we validate the login user and server authentication.

File Upload

In this module, after login the owner will upload using Block chain (hash functions) the file details and it'll be stored within the database.

Service Provider Login

In this module, given details to login and consider what are all the files uploaded by owner.

Request Accept and Reupload

Owner request a file to service provider which file want to re uploading give accept and owner reuploading a same file name and stored a database.

User Request

In this module, the user are going to be sending the file request to the third party that files, the user needs the access. Without the permission form the third party and repair provider, the user can't download the file.

Response from third and service provider sends key

In this module, the Third party send key and repair provider send key through user mail are going to be giving the acceptance to the user that file needs the access. After the acceptance, the file and cspkey are going to be sent to the user.

Download the file

In this module, after getting the key from the third party and service provider, the user can download the file using the key provided by the service provider.

VI. HOMOMORPHIC ENCRYPTION ALGORITHM

Homomorphic encryption is an encryption algorithm that's also a homomorphism. It allows the recipient of encrypted data to encrypt the results of some computation without knowing the inputs. the foremost popular example for the utilization of homomorphic encryption is where a knowledge owner wants to send data up to the cloud for processing, but doesn't trust a service provider with their data. employing a homomorphic encryption scheme, the info owner encrypts their data and sends it to the server. The server performs the relevant computations on the info without ever decrypting it and sends the encrypted results to the info owner. the info owner is that the just one ready to decrypt the results, since they alone have the key key. a way stronger and secure encryption than private and public key encryption developed is Homomorphic encryption. Homomorphic encryption may be a technique of encrypting the plaintext and performing computations on the encrypted text without disclosing the plaintext i.e. without decrypting it. Homomorphic Encryption are often called building blocks of recent cryptography because it is employed in many tools of cryptography.

VII. CONCLUSION

This paper proposes a blockchain-enabled deduplicatable data auditing mechanism to enhance the efficiency of the network storage service and protect the users' data. With the help of the deduplication technology, the network storage service provider can remove the duplicate data that the user outsourced and save just one copy, thereby reducing the storage burden. On this basis, we designed a blockchain-based data audit mechanism to make sure the integrity of outsourced data and avoid repeated audits by multiple tenants. The blockchain technique is introduced to record the info auditing log, thereby monitoring untrusted TPA during the data auditing process.

We think about utilizing a savvy agreement to actualize a blossom channel and an irregular number generator on the blockchain, to accomplish a programmed client-side information deduplication without the need for network capacity specialist co-op inclusion. Also, we tend to utilize the blockchain innovation to execute a completely decentralized information reviewing component without the need for a confided in TPA.

REFERENCES

[1] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 12, pp. 2231–2244, 2012.

[2] L. Zhou, A. Fu, S. Yu, M. Su, and B. Kuang, "Data integrity verification of the outsourced big data in the cloud environment: A survey," J. Netw. Comput. Appl., vol. 122, pp. 1–15, 2018..

[3] S. Li, C. Xu, and Y. Zhang, "CSED: Client-side encrypted deduplication scheme based on proofs of ownership for cloud storage," J. Inf. Security Appl., vol. 46, pp. 250–258, 2019.

[4] S. K. Madria, "Tutorial I: Security and privacy of big data in a cloud environment," 2015 11th International Conference on Innovations in Information Technology (IIT), Dubai, United Arab Emirates, pp XXXV-XXXVI ,2015

[5] Dharani P and Berlin M.A., "Survey on secret sharing scheme with deduplication in cloud computing," 2015 IEEE

9th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, pp 1-5 , 2015

[6] K. He, J. Chen, R. Du, Q. Wu, G. Xue and X. Zhang, "DeyPoS: Deduplicatable Dynamic Proof of Storage for Multi-User Environments," in IEEE Transactions on Computers, vol. 65, no. 12, pp. 3631-3645, 1 Dec. 2016

[7] W. Chen et al., "Cooperative and Distributed Computation Offloading for Blockchain-Empowered Industrial Internet of Things," in IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8433-8446, Oct. 2019

[8] Y. Yu, Y. Li, B. Yang, W. Susilo, G. Yang, and J. Bai, "Attribute-based cloud data integrity auditing for secure outsourced storage," IEEE Trans. Emerg. Topics Comput., vol. 14, no. 8, pp. 1–13, 2017.

[9] P. Puzio, R. Molva, M. Onen, and S. Loureiro, "Cloudedup: Secure ¨ deduplication with encrypted data for cloud storage," in Proc. 5th Int. Conf. Cloud Comput. Technol. Sci., vol. 1, 2013, pp. 363–370.

[10] J. Shen, D. Liu, M. Z. A. Bhuiyan, J. Shen, X. Sun, and A. Castiglione, "Secure verifiable database supporting efficient dynamic operations in cloud computing," IEEE Trans. Emerg. Topics Comput., pp. 1–11, 2017, doi: 10.1109/TETC.2017.2776402.

[11] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proc. 14th ACM Conf. Comput. Commun. Security, 2007, pp. 598–609.

[12] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in Proc. IEEE Conf. Commun. Netw. Security (CNS), 2013, pp. 145–153

[13] Q. Zheng and S. Xu, "Secure and efficient proof of storage with deduplication," in Proc. 2nd ACM Conf. Data Appl. Security Privacy, 2012, pp. 1–12.

[14] D. Reinsel, J. Gantz, and J. Rydning, "Data age 2025: The digitization of the world: from edge to core," Int. Data Corporation(IDC), pp. 1–28, 2018.

[15] F. Zafar, A. Khan, S. U. R. Malik, M. Ahmed, A. Anjum, M. I. Khan, N. Javed, M. Alam, and F. Jamil, "A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends," Computers & Security, vol. 65, pp. 29–49, 2017.

[16] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proc. 18th ACM Conf. Comput. Commun. Security, 2011, pp. 491–500.

[17] J. Gratz and D. Reinsel, "The digital universe decade-are you ready?" IDC White Paper, pp. 1–16, 2010.

[18] Y. Shin, D. Koo, and J. Hur, "A survey of secure data deduplication schemes for cloud storage systems," ACM Comput. Surveys (CSUR), vol. 49, no. 4, pp. 74–112, 2017.

[19] H. Shacham and B. Waters, "Compact proofs of retrievability," in Proc. Int. Conf. Theory Appl. Cryptol. Inf. Security, 2008, pp. 90–107.

★★★

INTER COMMON NEIGHBOR CONNECTIONS IN SUPERVISED LINK PREDICTION

[1]MANOJ KUMAR, [2]HARSHDEEP SINGH, [3]HIMANSHU SINGH RATHEE, [4]JATIN SOOD

[1,2,3,4] Department of Computer Science and Engineering, Delhi Technological University, Delhi-110042, India
E-mail: [1]mkumarg@dce.ac.in, [2]himanshusinghrathee_2k17co136@dtu.ac.in, [3]harshdeepsingh_2k17co130@dtu.ac.in, [4]jatinsood_2k17co148@dtu.ac.in

Abstract - Most social, biology, and information systems can use a network to define, therefore, examination of complex networks has become an important branch of many scientific fields. Guess the link to predict whether there will be a connection between the two nodes based on the relevant information and the available information of the link is the Link Prediction Problem. Recommender systems are the most common examples where link prediction is used extensively. The main approach to solve the Link Prediction Problem is using Supervised Learning; a score based approach where features from nodes of current snapshot of a graph are extracted and based on that future links are predicted. In this paper, we propose a supervised learning link prediction model using inter-common neighbor connection strength as a feature that considers how well the common neighbors of a node pair are connected alongside other commonly used graph topology based features to predict future links in social network based datasets. We compare how using inter-common neighbor connection strength improves the predictions using AUC-ROC score and F1 score as metric.

Keywords - Link Prediction, Recommender Systems, Supervised Learning, AUC-ROC score, F1 score

I. INTRODUCTION

In the midst of rapid growth and advancement, online social interacting has become a vital section of people's daily routine. Most social information systems can be represented through a network, where nodes represent individuals and margins represent human relationships or interactions between people. Therefore, examination of complex networks has become an important branch of many scientific fields. Guessing the link is an important function in linking mines. Guess the link to predict whether there will be a connection between the two nodes based on the relevant information and the available information of the link is the Link Prediction Problem. Link prediction application is not just restricted to social networking but can also be used in other fields. Applications in Bioinformatics are to detect interactions between proteins; Application in e-commerce, it is used to build the backbone of recommendation system; and Application in security sector, Detection of hidden criminal groups and gangs. Guess the link is closely related to most places. Therefore, in recent years many new approaches and integration algorithms have been developed to solve Link Prediction Problem.

There are plenty of similarity-based methods for link prediction in the literature. These methods essentially differ on what approach they use to estimate the similarity score between two nodes, which is then used to compute the likelihood of each non-existing link. Some methods estimate similarity based on neighborhood, i.e., they are based on local structural information, while other methods may consider paths of different length between the nodes to take semi-local information into account or may first need to traverse the whole graph for global structural information and then estimate the likelihood of non-existing links based on this information.

In this paper, we propose a new graph topology based metric to represent node-pair similarity called inter-common neighbor connection strength which is the total number of connections among the common neighbors. We make a generic link prediction model using the commonly used graph topology based features, namely Jaccard Index, Preferential Attachment, Common Neighbor, Ademic Ader and then creating a new model by adding the proposed metric as feature in the generic model for predicting future links. We then compare the predictions using auc-roc score and f1 score.

In this research, we have used XGBoost Classifier for predictions [1].

II. LINK PREDICTION PROBLEM

Liben-Nowell and Kleinberg have formalized the link prediction problem in the following way.
Let $G(V, L)$ be a network within the time period of $G[t, t_1]$ where V represents the set of nodes and L represents the set of links. For the next time period $G(t, t_2]$, the network might change. The link prediction focuses on how to predict the evolution of links, that is, how $L_{[t,t_1]}$ will differ from $L_{(t_1,t_2]}$. [2]

In simpler words, based on current graph topology, we predict most probable future links. It is a classification problem. In this research, we will just find the possibilities of addition of new links and not the removal of existing links in a network. Real-world social networks however are dynamic in nature and with time new links get added and old ones may get removed.

2.1. Graph topology based features

In the supervised learning model of basic link prediction, there are many matrices used to indicate certain features of a node pair which can be used for future link predictions. These metrics are based on graph topology. Throughout this section, $N(x)$ will represent the set of nodes in direct connection to node x and $N(y)$ will represents the set of nodes in direct connection to node y. $|S|$ represents the total number of values in a set S. Some of the commonly used prediction methods are discussed below:

COMMON NEIGHBOUR: Common Neighbor apprehends the idea of a node pair having more no of nodes as common connections are more likely to be connected in future than those who don't. It is the most common metric used when it comes to predicting future links [2, 3, 4]. It is calculated using the following method:

$$CN(x,y) = |N(x) \cap N(y)| \quad (1)$$

JACCARD INDEX: Jaccard Index improves the issue which is in Common Neighbor i.e. two nodes might have many neighbors out of which some are common but it doesn't mean they are more related than two nodes having less neighbors but more common neighbors. So in this we divide the common neighbors with the total number of neighbors to attain normalization for comparison [2]. It is calculated using the following method:

$$J(x,y) = \frac{|N(x) \cap N(y)|}{|N(x) \cup N(y)|} \quad (2)$$

COSINE SIMILARITY: It is also used to measure similarity among node pairs. In this method the similarity is calculated based on the dot product of two vectors [5]. It is calculated by following formula:

$$CS(x,y) = \frac{|N(x) \cap N(y)|}{\sqrt[2]{|N(x)| \times |N(y)|}} \quad (3)$$

PREFERENTIAL ATTACHMENT INDEX: In preferential attachment index the closeness of two nodes is defined the product of the degree of the two nodes. It works on the principal that a node which has more connections today will most likely make new connections in the future [6]. A higher value represents the higher closeness of the nodes. It is calculated using the following method:

$$PA(x,y) = |N(x)| \times |N(y)| \quad (4)$$

ADAMIC ADAR INDEX: It is also a Node based similarity index similar to common neighbor index above. It is defined by the reciprocal of the log of the degree of the shared neighbors of the node x and y. It is based on the idea that lesser the degree of common nodes corresponds to better chances of connection of two nodes [7]. It is calculated by the following method:

$$AA(x,y) = \sum_{U \in N(x) \cap N(y)} log(|N(u)|) \quad (5)$$

KATZ COEFFICIENT: This method takes all the number of distinct lengths of paths between two nodes [8]. The number of paths between node x and node y in length l is stated as $|path_{xy}^{<l>}|$ is calculated and multiplied by factor β^l. By summation of all the results for all possible distinct values of l, we get the Katz coefficient value for a node pair. In [9] it is discussed that in most cases Katz coefficient is the best metric for node-pair similarity. However, it is computationally very expensive to calculate for each and every node-pair.

$$KC(x,y) = \sum_{l=1}^{\infty} \beta^l . |path_{xy}^{<l>}| \quad (6)$$

INTER COMMON NEIGHBOUR CONNECTIONS: We propose this new method which calculates the total number of connections among the common neighbors. The better the common neighbors are connected among themselves, higher the probability that the node pair will connect in future. It can be represented by:

$$ICNC(x\,y) = \sum_{u \in |N(x) \cap N(y)|} \sum_{v \in |N(x) \cap N(y)|} \frac{has\_edge(u,v)}{2} \quad (7)$$

Where has_edge(u, v) = 1 if there is an edge between u, v. Otherwise it is 0. Also, the datasets we have used does not contain self-loops.

2.2. Performance Metric

In Machine Learning it is important to do performance evaluation for an algorithm. As we already discussed, most of the real world network models are sparse in nature. The ratio of linked edges to non-linked edges is less than 0.001. In such cases using accuracy as a prediction metric does not make sense, because even if we create a model which predicts that links are never going to be made in future, accuracy would come out to be more than 99%, but, clearly, this is not the best prediction model.

We will use the area under the curve value for receiver operator characteristic (AUC-ROC) curve for evaluation [10]. AUC-ROC score is 0.5 in case of independent and identical distribution, i.e. random prediction. A value greater than 0.5 shows us how much better our model is performing when compared to a model which does random predictions.

The second performance we are going to use is the F1 score. F1 score is the harmonic mean of precision and recall. Precision tells us chances that we predict a link would form and it actually forms in future. Recall tells us what proportion of links that we identified correctly will be formed in future.

2.3. Datasets

Multiple open source datasets are available for the link prediction problem. We will be using 3 social

network datasets. The datasets used are:

FACEBOOK DATASET: This data was collected from survey participants using a Facebook application [11].
Nodes: 4039
Edges: 88234

FACEBOOK PAGES DATASET: This dataset represent blue verified Facebook page networks of different categories where nodes represent the pages and edges are mutual likes among them [12].
Nodes: 3892
Edges: 17262

HAMSTERSTER DATASET: This dataset contains friendship and family links of the users of the Hamsterster social media platform [12].
Nodes: 2426
Edges: 16630

III. DATA PREPROCESSING

The data is present in .txt file having two columns denoting nodes having a link. We create a pandas data frame with all possible node pairs and a link attribute which is 1 if there is a link between the node pair, else it is 0. For each node pair, we calculate the common neighbors, preferential attachment, adamic-adar and jaccard coefficient value.
We then split this data into training and testing data. Since the data frame is sparse in nature, we ensure equal ratios of linked node pairs to non-linked node pairs are used for training and testing. For the sake of this experiment, we have used XGBoost Model to predict results and get AUC-ROC score and F1 score.

IV. EXPERIMENTAL RESULTS

The results found by experimenting on datasets on 3 different test runs are shown below. The results show that using the proposed feature alongside other features discussed almost always improves the prediction results.

FACEBOOK DATASET

	Test Run 1		**Test Run 2**		**Test Run 3**	
	w/o proposed feature	**with proposed feature**	**w/o proposed feature**	**with proposed feature**	**w/o proposed feature**	**with proposed feature**
AUC-ROC score	0.845	0.849	0.848	0.848	0.846	0.849
F1 score	0.736	0.741	0.740	0.740	0.737	0.741

Table 1: AUC-ROC scores and F1 scores for Facebook dataset using and without using the inter-common neighbor connections as a feature in 3 different test runs

FACEBOOK PAGES DATASET

	Test Run 1		**Test Run 2**		**Test Run 3**	
	w/o proposed feature	**with proposed feature**	**w/o proposed feature**	**with proposed feature**	**w/o proposed feature**	**with proposed feature**
AUC-ROC score	0.783	0.804	0.785	0.796	0.783	0.794
F1 score	0.668	0.698	0.674	0.695	0.668	0.692

Table 2: AUC-ROC scores and F1 scores for Facebook Pages dataset using and without using the inter-common neighbor connections as a feature in 3 different test runs

HAMSTERSTER DATASET

	Test Run 1		**Test Run 2**		**Test Run 3**	
	w/o proposed feature	**with proposed feature**	**w/o proposed feature**	**with proposed feature**	**w/o proposed feature**	**with proposed feature**
AUC-ROC score	0.761	0.791	0.770	0.789	0.753	0.789
F1 score	0.634	0.691	0.654	0.687	0.624	0.687

Table 3: AUC-ROC scores and F1 scores for Hamsterster dataset using and without using the inter-common neighbor connections as a feature in 3 different test runs

V. CONCLUSION AND FUTURE SCOPE

After reviewing the results on different social media datasets with and without using the inter-common neighbor connections as a feature, we can say that in most cases, using inter-common neighbor connections as a feature in the supervised learning model for link prediction improves the predictions with up to 3% increase in the AUC-ROC score.

Table 1 which shows the AUC-ROC score and F1 score for Facebook dataset shows that using intercommon neighbor connections has very little effect on the prediction results. Table 2 and Table 3 however shows that inter-common neighbor connection makes the predictions noticeably better.

In our research, we have not used any path based feature such as katz coefficient due to their computational expenses. In [13, 14], the author has discussed random walk based approach for shortest path calculation which is much less computationally expensive and shows great results in improving predictions.

In [15], problems with the current link prediction approach are discussed. Social Networks are dynamic n nature were links are created and deleted over time. Hence, predicting links based on a snapshot of a graph is technically not the most effective way.

REFERENCES

[1] C. Tianqi and G. Carlos, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.

[2] D. Liben-Nowel and J. Kleinberg, "The link prediction problem for social networks," in roceedings of the 12th ACM International Conference on Information and Knowledge Management (CIKM '03), New York, 2003.

[3] W. Cukierski, B. Hamner and B. Yang, "Graph-based features for supervised link prediction," in The 2011 International Joint Conference on Neural Networks, California, 2011.

[4] L. Lü, C.-H. Jin and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," Physical review. E, Statistical, nonlinear, and soft matter physics, vol. 80, 2009.

[5] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications, vol. 390, no. 6, pp. 1150-1170, 2011.

[6] M. Newman, "Clustering and preferential attachment in growing networks," Physical review. E, Statistical, nonlinear, and soft matter physics, vol. 64, 2001.

[7] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," Social Networks, vol. 25, no. 3, pp. 211-230, 2003.

[8] L. Katz, "A new status index derived from sociometric analysis," Psychometrika, vol. 18, no. 1, pp. 39-43, 1953.

[9] F. Gao, K. Musial, C. Cooper and S. Tsoka, "Link Prediction Methods and Their Accuracy for Different Social Networks and Network Metrics," Scientific Programming, vol. 2015, 2015.

[10] J. A. Hanley and B. Mcneil, "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve," Radiology, vol. 143, pp. 29-36, 1982.

[11] J. McAuley and J. Leskovec, "Learning to Discover Social Circles in Ego Networks," in Neural Information Processing Systems (NIPS), 2012, Nevada, 2012.

[12] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, p. 4292–4293, 2015.

[13] W. Liu and L. Lü, "Link Prediction Based on Local Random Walk," Europhysic Letter, vol. 89, 2010.

[14] L. Backstrom and J. Leskovec, "Supervised Random Walks: Predicting and Recommending Links in Social Networks," Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011, pp. 635-644, 2011.

[15] R. R. Junuthula, K. S. Xu and V. K. Devabhaktuni, "Evaluating Link Prediction Accuracy in Dynamic Networks with Added and Removed Edges," 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom, pp. 377-384, 2016.

★★★

INHIBITION OF SARS-COV-2 PROTEIN BY BIOACTIVE COMPOUNDS OF EDIBLE MUSHROOM; A BIOINFORMATICS INSIGHT

[1]DEBANJAN MITRA, [2]PRADEEP K. DAS MOHAPATRA

[1,2]Department of Microbiology, Raiganj University, Raiganj, WB, 733134
Email: [1]debanjanmitra267@gmail.com, [2]pkdmvu@gmail.com

Abstract - Nowadays most of the researches of world are going to find a remedy of COVID-19. Still, WHO doesn't completely recommended a drug that will surely treat COVID-19. Although several country use hydroxychloroquine and remdesivir, but they don't show any remarkable effect on treatment of COVID-19. This present study has proposed a drug that will inhibit multiple proteins of coronavirus. Here, we have analyzed some bioactive drug compounds which are already approved for their therapeutic effect. Those compounds are found in edible mushroom. They are identified by ADMET properties which show their non-toxic nature and high binding energies with targeted proteins. All those bioactive drug compounds are passing through many insilico processes and fulfil all the parameters. By this study, Polyozellin appears as a best bioactive drug compounds which will inhibit multiple proteins of coronavirus and used as oral drug. Bioactive compounds are easily used as medicine for their low cost and they don't have any side effects. So, we proposed those bioactive compounds which will be used as a drug medicine in treatment of COVID-19.

Keywords - COVID-19, Therapeutic effect, ADMET properties, Oral drug

I. INTRODUCTION

The novel coronavirus/SARS-CoV-2, which is a nightmare for the people of world, was first reported in late December, 2019 in Wuhan city at China [1-3]. Novel corona (2019-nCoV) has a close relationship with Middle East Respiratory Syndrome Coronavirus (MERS-CoV) outbreaks [4]. Although it was started from China, but now there is a terrible outbreak of this disease all over the world. It triggered the scientific community to rapidly development of testing kits and cure for infected patients. The mortality rate is increase day by day but no cure has been found yet. On March 11, 2020 the World Health Organization (WHO) declared the COVID-19 outbreak as a pandemic.

It is a member of β-coronavirus family. Generally it contain four types of protein i.e. nucleocapsid protein, spike protein, membrane protein and ORF protein. Nucleocapsid proteins of SARS-CoV-2 help in packaging the positive strand viral genome RNA into a helical ribonucleocapsid. It also helps in virion assembly through its interactions with the viral genome. 3C-like protease is one type of ORF polyprotein, which helps in cleaving the C-terminus of replicase polyprotein at 11 sites [5]. RNA replicase is also an ORF polyprotein. It participates in viral replication by acting as an ssRNA-binding protein [6]. The 3D structures of these proteins have been resolved by X-ray crystallography [7]. Many *in silico* or computational studies have utilized these proteins as the main targets for drug screening and repurposing from various sources [8-11].

Natural bioactive molecules have now made their special manifestations as a drug. Because they can be used very easily in human body and they do not have any harmful effect. Generally all these biomolecules are present in our food and can easily enter in our body. Some researchers have suggested the use of bioactive molecules to prevent COVID 19[12-17]. Myricetin is a member of polyphenolic compounds, with antioxidant properties have been found in some wild edible mushroom (*Cantharellus cibarius, Craterellus cornucopioides* etc.) [18]. It is already used as a drug [19]. Polyozellin and thelephoric acid are a key components of an edible mushroom *Polyozellus multiplex* which cure stomach cancer by inhibiting prolyl endopeptidase[20-21]. Clavilactone is a key component of *Ampulloclitocybe clavipes [22] and an inhibitor of protein tyrosin kinase.* Dibenzofuran is a heterocyclic organic compound which is found in *Boletopsis leucomelas*. It is used as a precursor of an anti-inflammatory drug, furobufen [23].

We attempted docking of 5 compounds (myricetin, polyozellin, thelephoric acid, clavilactone and dibenzofuran) which are found in some edible mushroom with 3D structure of SARS-CoV-2 proteins which are available in database. Here we took 3 proteins i.e. 3C like main protease, RNA polymerase and nucleocapsid protein as a target which play important roles in SARS-CoV-2. These 5 compounds were subjected to quantum mechanical studies and QSAR analysis to detect their activity as a drug. The results from these *in silico* binding studies are reported. It is possible that these compounds may inhibit those proteins of SARS-CoV-2.

II. MATERIALS AND METHODS

A. Target protein structures retrieval

The target protein was extracted from RCSB PDB [24] database. Those 3D structures are selected by their high resolution. Here we chose 3 types of

proteins i.e. 3C like main protease (7BRO), RNA replicase (6WXD) and nucleocapsid protein (7C22) from coronavirus. So the effect of drug molecule on multiple coronavirus proteins will be widely examined.

B. Drug compound retrieval

Those bioactive compounds i.e. myricetin, polyozellin, thelephoric acid, clavilactone and dibenzofuran from difference fungal sources were extracted from PUBCHEM [25] database in .sdf format.

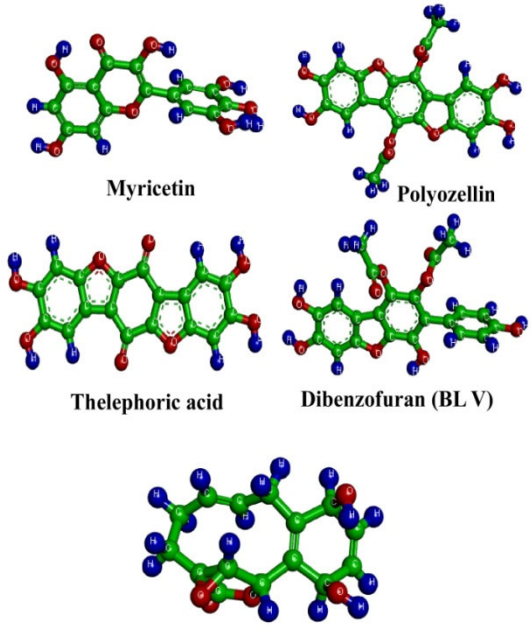


Figure 1: 3D structure of myricetin, polyozellin, thelephoric acid, dibenzofuran (BL V) and clavilactone A

C. Surface area of protein

To identify the pockets of 3C like main protease (7BRO), RNA replicase (6WXD) and nucleocapsid protein (7C22) from coronavirus, the CASTp server was used. The pocket indicates the accessible area of proteins.

D. Drug likeness with descriptor properties

The drug likeness of Myricetin, Polyozellin, Thelephoric Acid, Clavilactone and Dibenzofuran was evaluated by the Lipinski rule of five [26]. This rule is important for drug pharmacokinetics in the human body and provides information regarding the application of the ligands as drugs. This rule is only applicable to oral drug.

E. Protein ligand complex preparation

The ligands were minimized in semi-empirical process by PM3 using VEGA ZZ software [27]. The protein ligand complex was created by PyMol software [28].

F. Molecular docking studies on proteins

Autodock 4.2 [29] of the Scripps Research Institute was used for construction of ligand and docking on selected proteins i.e. 3C like main protease, RNA polymerase and nucleocapsid protein, respectively. All target proteins were docked by this software. The site of the target intended for docking was placed in a grid box with conventional dimension and grid spacing. Grid space was used as previous research reports [30]. Root of ligand also ascertain by Autodock 4.2. Rotatable bonds were established and nonpolar hydrogen atoms were integrated.

III. RESULT AND DISCUSSIONS

A. Drug properties of selected bioactive compounds

Lipinski rule of five was applied to evaluate descriptor properties which indicate those bioactive compounds as a drug. According to Lipinski rule of five, LogP or Octanol-water partition coefficient of a drug compound will be in lower than +5, molecular weight of a compound will be ≤500, HBA (hydrogen bond acceptor) of a compound will be ≤10, HBD (hydrogen bond donor) of a compound will be ≤5. Here also some additional important descriptor properties like total polar surface area (PSA), number of atoms (natoms), number of heavy atoms, volume, ovality, molar refractivity, Devis HLB are also calculated.

Fungal bioactive	Myricetin	Polyozellin	Thelephoric acid	Dibenzofuran (BL V)	Clavilactone A
logP	1.42	1.67	0.20	2.68	-0.50
Molecular weight (Grams/mol)	318.24	438.34	352.25	424.36	276.29
HBD	6	4	4	4	2
HBA	8	10	8	9	5
PSA (Å²)	138	160	141	147	79.3
Total Atoms	33	46	34	47	36
Heavy atoms	23	32	26	31	20
Volume (Å³)	246.10	339.10	267.10	346.10	242.10
Ovality	1.55	1.73	1.36	1.71	1.51
Molar refractivit	76.47	106.86	88.59	105.90	70.98
Davies' HLB	17.80	21.75	15.30	19.50	11.18

Table 1- Bioactive compounds with logP, molecular weight, HBD, HBA, PSA, total atoms, heavy atoms, volume, ovality, molar refractivity and Davies' HLB

Values of all the parameters follow Lipinski rule of five. Logp indicates the lipophilicity of a drug compound [31]. Lipophilicity is a crucial dictating aspect in a compound's absorption, dispersal in the body, puncturing across vital membranes and biological barriers, metabolism and excretion. The highest positive value of Dibenzofuran (BL V) indicates that it has highest lipophilicity [32]. Generally, lower than 500 Dalton molecular weight compounds are used as oral drugs [33] and all those

bioactive compounds are indicating that they can be used in the same way. All those compounds don't have large number of HBA and HBD. If a drug interacts too dynamically with the water in the body, it will have an effect on the absorption, meaning it will not get to the essential target in the body [34]. Polar surface area (PSA) is predicting the intestinal absorption of drugs in humans [35]. It is high in Polyozellin and low in Clavilactone A. Number of heavy atoms is high in Polyozellin and low in Clavilactone A. It shows a direct correlation with PSA. Ovality of the molecule is highly significantly directly correlated with molecular weight. Increasing the values of ovality would result to a lower log IC50 [36]. Previous reports suggest that the molar refractivity of a compound between 40 to 130 could be used in the development of druglike chemical libraries [3-387]. This value of each compound is in the above specific range. A *high HLB* value indicates *high* water-solubility. If the drug's *HLB* value is *higher* than 10, it is more hydrophilic [39]. *High HLB* helps to increase the permeability of the *drug [40] that means* Polyozellin has higher permeability.

B. Surface area and volume of target proteins

Nucleocapsid protein (7C22) shows highest pocket area which is 1580.25 A^2 and the volume is 3033.76 A^3. 3C like main protease (7BRO) shows 518.26 A^2 as pocket area and 394.22 A^3 as total volume. RNA replicase (6WXD) shows lowest pocket area which is 574.25 A^2 and the volume is 329.25 A^3. Large surface area has a greater chance to participate with ligand molecules. Generally pocket areas are occupied by hydrophobic residues which have a tendency to increase the protein-ligand complex stability. So, those proteins have higher hydrophobic residues might shows strong binding energies with ligands.

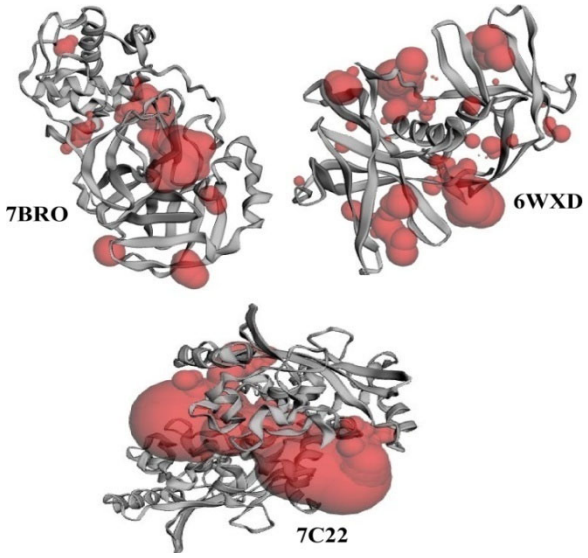


Figure 2: 3D structure of 3C like main protease (7BRO), RNA replicase (6WXD) and nucleocapsid protein (7C22) with their pocket area (red) for drug binding

C. Identification of best drug by molecular docking

Molecular docking studies shows the binding energies which are indicating that those drug is suitable for that protein or not. Here we want to identify a single drug that will effect on not only a single protein; it has an ability to inhibit multiple proteins of coronavirus. So, that drug will easily stop COVID-19.

Table 2- Drugs name with their PUBCHEM ID and binding energies with 3C like main protease (7BRO), RNA replicase (6WXD) and nucleocapsid protein (7C22) from coronavirus

In case of 3C like main protease, Myricetin shows highest binding energies followed by Polyzellin, Dibenzofuran, Thelephoric acid and Clavilactone A. In case of RNA replicase, Dibenzofuran shows highest binding energies that indicates its inhibitory effect. In nucleocapsid protein docking analysis, Polyozellin appears as a best drug to inhibit it between all those drug candidates. In case of multiple protein inhibition by a single drug, Polyozellin appears as best followed by Dibenzofuran, Thelephoric acid, Myricetin and Clavilactone A. Myricetin binds with 3C like main protease by 3 hydrogen bonds with threonine and isolusine, 1 pi-cation with aspartic acid and 1 pi-anion bond with arginine and 8 van der wall interactions. Dibenzofuran binds with RNA replicase by 4 pi-alkyl bonds with valine and isolusine, 1 pi-sigma bond with isolusine, 1 hydrogen bond with serin and 5 van der wall interactions. Polyozellin binds with nucleocapsid protein by 6 pi-alkyl bonds with alanine, 1 pi-pi interaction with tryptophan, 2 pi donar hydrogen bond with tryptophan and alanine, 1 carbon-hydrogen bond with proline, 1 conventional hydrogen bond with proline, 11 van der wall interactions.

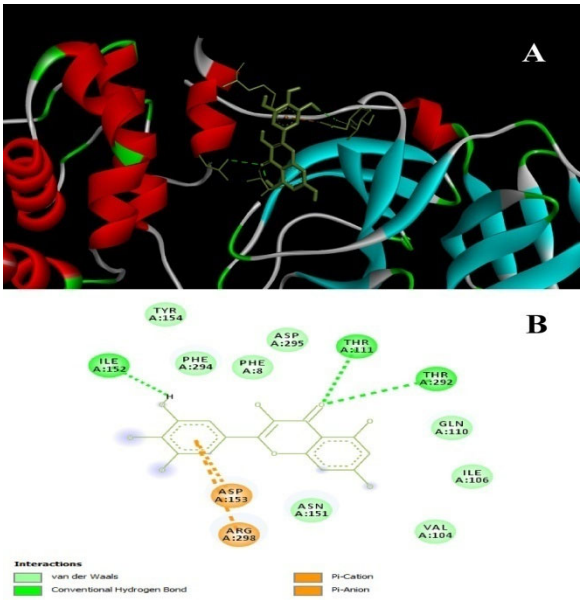


Figure 3: A. 3D interaction of Myricetin with 3CL main protease (7BRO) B. 2D interaction of Myricetin with 3CL main protease (7BRO) shows its bonding with amino acids

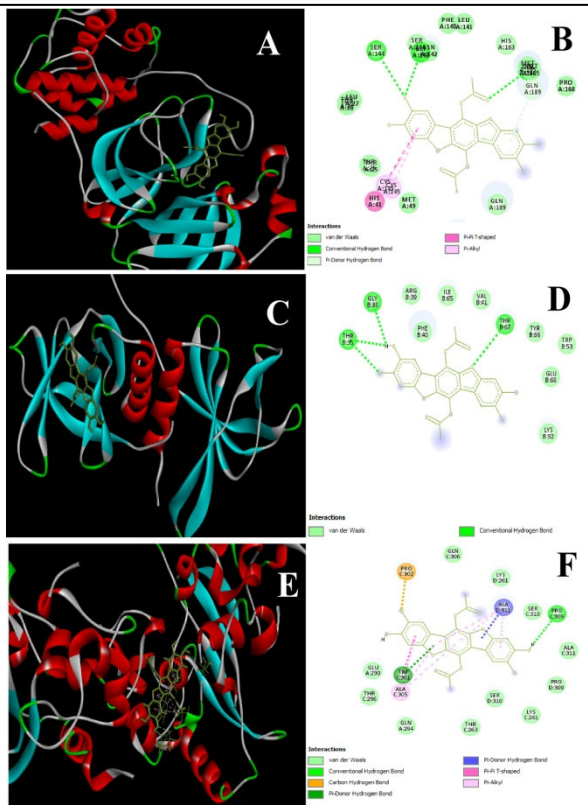


Figure 4: A. 3D interaction of Polyozellin with 3CL main protease (7BRO) B. 2D interaction of Myricetin with 3CL main protease (7BRO) shows its bonding with amino acids C. 3D interaction of Polyozellin with RNA replicase (6WXD) D. 2D interaction of Myricetin with RNA replicase (6WXD) shows its bonding with amino acids E. 3D interaction of Polyozellin with nucleocapsid protein (7C22) D. 2D interaction of Myricetin with nucleocapsid protein (7C22) shows its bonding with amino acids.

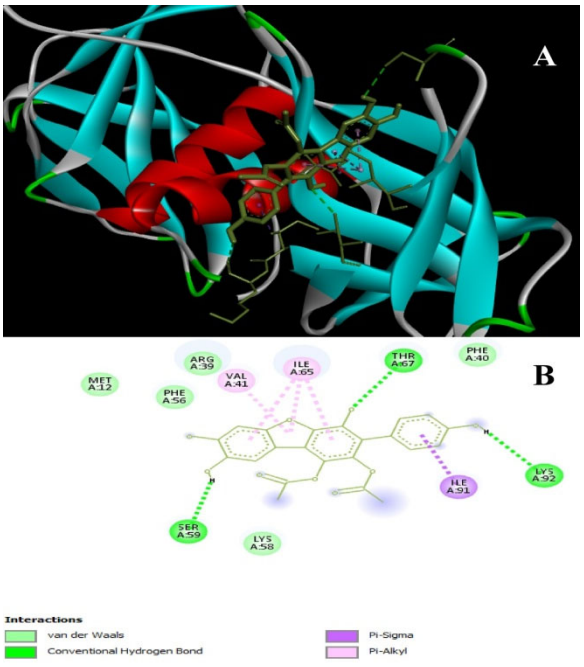


Figure 5: A. 3D interaction of Dibenzofuran (BL V) with nucleocapsid protein (7C22) B. 2D interaction of Dibenzofuran (BL V) with nucleocapsid protein (7C22) shows its bonding with amino acids.

IV. CONCLUSION

Present study shows how some bioactive compounds will used as drug in medicine in treatment of COVID-19. Multiple types of protein from coronavirus taken as target to find a best drug that will inhibit multiple types of protein and stop COVID-19. Myricetin, Polyozellin, Thelephoric Acid, Clavilactone and Dibenzofuran are shows their ability to used as an oral drugs as they fulfill the Lipinski rule of five. High binding energies of those compounds with target proteins indicate their strong binding. Between all those bioactive drug compounds Polyozellin appears as a drug which can inhibit multiple proteins of coronavirus. So, those bioactive drug compounds will be used in treatment of COVID-19 due to their safe, low cost, non toxic nature and high binding energies.

REFERENCES

[1] A. Hasan, B.A. Paray, A. Hussain, F.A. Qadir, F. Attar, F.M. Aziz, M. Sharifi, H. Derakhshankhah, B. Rasti, M. Mehrabi, K. Shahpasand, "A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin", J Biomol. Struct. Dyn. pp. 1-9, April 2020.

[2] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding", The Lancet. Vol. 395, pp. 565-574, February 2020.

[3] S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G.Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, "A new coronavirus associated with human respiratory disease in China", Nature, Vol. 579, pp. 265-269, February 2020.

[4] Q. Han, Q. Lin, S. Jin, L. You, "Recent insights into 2019-nCoV: a brief but comprehensive review", J Infec. Vol. 25 February 2020.

[5] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, "Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors", Science, Vol. 368, pp. 409-412, April 2020.

[6] D.R. Littler, B.S. Gully, R.N. Colson, J. Rossjohn, "Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9", Iscience, Vol. 23, PP. 101258, June 2020.

[7] R. Zhou, R. Zeng, A. von Brunn, J. Lei, "Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein", Mol. Biomed. Vol. 1, pp. 1-11.

[8] S. Khaerunnisa, H. Kurniawan, R. Awaluddin, S. Suhartati, S. Soetjipto, "Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study", Prepr. Vol. 13, pp. 1-4, March 2020.

[9] B. Shah, P. Modi, S.R. Sagar, "In silico studies on therapeutic agents for COVID-19: Drug repurposing approach", Life Sciences, Vol. 9, pp. 117652, April 2020.

[10] S. Singh, M.F. Sk, A. Sonawane, P. Kar, S. Sadhukhan, "Plant-derived natural polyphenols as potential antiviral drugs against SARS-CoV-2 via RNA-dependent RNA polymerase (RdRp) inhibition: An in-silico analysis", J Biomol. Struct. Dyn. Vol. 23, pp. 1-6, July 2020.

[11] V. Chenthamarakshan, P. Das, I. Padhi, H. Strobelt, K.W. Lim, B. Hoover, S.C. Hoffman, A. Mojsilovic, "Target-specific and selective drug design for covid-19 using deep generative models", arXiv preprint arXiv, April 2004.

[12] S. Borkotoky, M. Banerjee, "A computational prediction of SARS-CoV-2 structural protein inhibitors from Azadirachta

indica (Neem)", J Biomol. Struct. Dyn. Vol. 26, pp. 1-7, May 2020.
[13] R. Islam, M.R. Parves, A.S. Paul, N. Uddin, M.S. Rahman, A.A. Mamun M.N. Hossain, M.A Ali, M.A. Halim, "A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2", J Biomol. Struct. Dyn. Vol. 7, PP. 1-2, May 2020.
[14] V.K. Bhardwaj, R. Singh, J. Sharma, V. Rajendran, R. Purohit, S. Kumar, "Identification of bioactive molecules from Tea plant as SARS-CoV-2 main protease inhibitors", J Biomol. Struct. Dyn. Vol. 8, pp. 1-3, May 2020.
[15] A. Kumar, G. Choudhir, S.K. Shukla, M. Sharma, P. Tyagi, A. Bhushan, M. Rathore, "Identification of phytochemical inhibitors against main protease of COVID-19 using molecular modeling approaches", J Biomol. Struct. Dyn. Vol. 2, pp. 01-21, May 2020.
[16] M.K. Gupta, S. Vemula, R. Donde, G. Gouda, L. Behera, R. Vadde, "In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel", J Biomol. Struct. Dyn. Vol. 13, pp. 1-11, April 2020.
[17] A.A. Elfiky, "Natural products may interfere with SARS-CoV-2 attachment to the host cell", J Biomol. Struct. Dyn. Vol. 5, pp. 1-10, May 2020.
[18] I. Palacios, M. Lozano, C. Moro, M. D'arrigo, M.A. Rostagno, J.A. Martínez, A. García-Lafuente, E. Guillamón, A. Villares, "Antioxidant properties of phenolic compounds occurring in edible mushrooms", Food Chem. Vol. 128, pp. 674-678, October 2011.
[19] J. Qian, H. Meng, L. Xin, M. Xia, H. Shen, G. Li, Y. Xie, "Self-nanoemulsifying drug delivery systems of myricetin: Formulation development, characterization, and in vitro and in vivo evaluation", Colloids and Surfaces B: Biointerfaces, Vol. 160, pp. 101-109, December 2017.
[20] E.J. Yang, K.S. Song, "Polyozellin, a key constituent of the edible mushroom Polyozellus multiplex, attenuates glutamate-induced mouse hippocampal neuronal HT22 cell death. Food & Func. Vol. 6, pp. 3678-3686, May 2015.
[21] I. Doskocil, J. Havlik, R. Verlotta, J. Tauchen, L. Vesela, K. Macakova, L. Opletal, L. Kokoska, V. Rada, "In vitro immunomodulatory activity, cytotoxicity and chemistry of some central European polypores", Pharma. Bio. Vol. 54, pp. 2369-2376, November 2016.
[22] Z.M Thu, K.K. Myo, H.T. Aung, M. Clericuzio, C. Armijos, G. Vidari, "Bioactive phytochemical constituents of wild edible mushrooms from southeast Asia", Molecules, Vol. 25, pp. 1972, January 2020.
[23] R.R. Martel, J.G. Rochefort, J. Klicius, T.A. Dobson, "Anti-inflammatory properties of furobufen", Can. J. Physiol. Pharmacol. Vol. 52, pp. 669-673, June 1974.
[24] A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P.E. Bourne, H.M. Berman, "The RCSB PDB information portal for structural genomics", Nucleic acids Res. Vol. 34, pp. 302-305, January 2006.
[25] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, "PubChem substance and compound databases", Nucleic acids Res. Vol. 44, pp. D1202-D1213, January 2016.
[26] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings", Adv. Drug Deliv. Rev. Vol. 23, pp. 3-25, January 1997.
[27] A. Pedretti, A. Mazzolari, G. Vistoli, "VEGA ZZ: a versatile toolkit for drug design and protein modeling", J Comput. Chem. Vol. 25, pp. 1605-1612, 2004.
[28] W.L. DeLano, "Pymol: An open-source molecular graphics tool", CCP4 Newsletter on protein crystallography, Vol. 40, pp. 82-92, March 2002.
[29] G.M. Morris, R. Huey, A.J. Olson, "Using autodock for ligand-receptor docking" Curr. Prot. Bioinfo. Vol. 24, pp. 8-14, December 2008.
[30] I. Ansary, H. Roy, A. Das, D. Mitra, A.K. Bandyopadhyay, "Regioselective synthesis, molecular descriptors of (1, 5-disubstituted 1, 2, 3-triazolyl) coumarin/quinolone derivatives and their docking studies against cancer targets", ChemistrySelect, Vol. 4, pp. 3486-3494, March 2019.
[31] D. Mitra, "Quantum mechanical descriptors of nilotinib's impurities", Quantum, Vol. 6, March 2018.
[32] G. Camenisch, J. Alsenz, H. van de Waterbeemd, G. Folkers, "Estimation of permeability by passive diffusion through Caco-2 cell monolayers using the drugs' lipophilicity and molecular weight", Eur. J. Pharm. Sci. Vol. 6, pp. 313-319, October 1998.
[33] J.D. Bos, M.M. Meinardi, "The 500 Dalton rule for the skin penetration of chemical compounds and drugs", Exp. Dermatol, Vol. 9, pp. 165-169, June 2000.
[34] I. Gashaw, P. Ellinghaus, A. Sommer, K. Asadullah, "What makes a good drug target?", Drug Discov. Today, Vol. 16, pp. 23-24, December 2011.
[35] D.E. Clark, "What has polar surface area ever done for drug discovery?", Futu. Med. Chem. Vol. 3, PP. 469-484, March 2011.
[36] D.E. Sumalapao, J.I. Janairo, N.G. Gloriani, "Dipole moment, solvation energy, and ovality account for the variations in the biological activity of HIV-1 reverse transcriptase inhibitor fragments", Anna. Res. Rev. Bio. January 2018.
[37] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, "A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases", J Combi. Chem. Vol. 1, pp. 55-68. January 1999.
[38] G.R. Bickerton, G.V. Paolini, J. Besnard, S. Muresan, A.L. Hopkins, "Quantifying the chemical beauty of drugs", Nature Chem. Vol. 4, pp. 90-98, February 2012.
[39] L. Wang, J. Dong, J. Chen, J. Eastoe, X. Li, "Design and optimization of a new self-nanoemulsifying drug delivery system", J Coll. Interf. Sci. Vol. 330, pp. 443-448, February 2009.
[40] A.K. Gurram, P.B. Deshpande, S.S Kar, U.Y. Nayak, N. Udupa, M.S. Reddy, "Role of components in the formation of self-microemulsifying drug delivery systems", Indian J Pharm. Sci. Vol. 77, pp. 249, May 2015.

★★★

PRE-ACCIDENTS DETECTION AND PREVENTION DEVICE USING ULTRASONICS, ACTI-GRAPH AND SPEED SENSORS

[1]BUELA PROMODINI, [2]KALYANI, [3]KESAVA, [4]ASHISH PREETHAM, [5]PREETHI, [6]PAVN KUMAR

[1]Assistant Professor (ECE), Dr. APJ Abdul Kalam IIIT ONGOLE, RGUKT-AP
[2]Student(ECE), Dr.APJ Abdul Kalam IIIT ONGOLE,RGUKT-AP
[3]Student(ECE), Dr.APJ Abdul Kalam IIITONGOLE,RGUKT-AP
[4]Student(ECE), Dr.APJ Abdul Kalam IIITONGOLE,RGUKT-AP
[5]Student(CSE)APJ Abdul Kalam IIITONGOLE, RGUKT-AP
[6]Student(MECH)APJ Abdul Kalam IIITONGOLE, RGUKT-AP
E-mail:[1]buelapromodini@gmail.com,[2]kalyani1850@gmail.com, [3]kesavaontipuli567@gmail.com,
[4]ashishpreethamturaka@gmail.com, [5]preethims1828@gmail.com, [6]ptaduvai@gmail.com

Abstract: This article addresses about designing of an electronic device to set up in cars that co-ordinates Ultra-sonic,Acti-graph,in-built speed sensors for avoiding accidents and detection and prevention.Some theoretical expressions were derived to run this device effectively.Under these expressions we got satisfactory results during prototype execution. On research,major accidents provoking conditions in India like crashing vehicles,rash and restless driving which leads to even death sometimes.So far on considering these above mentioned conditions,this device assures with maximum safety approach.

Keywords - Ultra-sonic Sensors, Acti-graph Sensors, Pre-accidents, Prototype Execution, Rash and Restless driving.

I. INTRODUCTION

Due to high population rate India is facing many provoking situations in that, one of the major complication is “ACCIEDENTS”.To overcome this issue we had a précised note point i.e. ‘A device which prevents the accidents’.

The device which detects the vehicle i.e., “Vehicle Detector” through the “Ultra Sonic Sensors”, controls the speed of the vehicle through the “Speed lock” & detects the condition of the driver i.e. “Drowsiness Detection” through the “Eye Blink Sensor. The complete vision of these three sensor based devices gives us a “MONITORING DEVICE”.
In this device firstly is about the ,Vehicle Detection by the usage of ‘ultrasonic sensors’ which means an electronic device that measures the distance of a target object emitting ultrasonic sound waves and converts the reflected sound in an electrical signal.Through this information we detect the vehicle. Then next through the speed lock system we detect the speed of the vehicle and control the speed as per the instructions. Lastone, is detecting the driver of the vehicle and checks the condition of the driver through the ‘eye blink sensor’ which gives complete information of the driver by illuminating the eye and eyelid area with infrared light. Alternate work of this devices in a single device gives as the output,which prevents accidents.

II. SALIENT FEATURES OF DEVICE

Vehicle detection,speed lock and drowsiness detection are the main features that are to be carried by the externally taken micro controller which furthers connected to ECU.These features can be implemented by taking inputs by three main sensors like advanced Ultrasonics,eye blink sensors and vehicle speed sensors.

Firstly to the car Ultrasonics are fixed in such a way that to detect any approaching from all angles. Ultrasonics gives you the idea of distance between owners’ vehicle and barrier or vehicle by constantly emitting sonic waves. As we know that rate of change of distance with respect to time speed of vehicle can be determined.

We know that Speed of an engine depends upon the fuel flow rate. Different cars have different fuel flow controlling devices such as carburetion and throttle for petrol engines and for diesel engines governors and fuel injectors. Now days electronic components are used in controlling speed of the engine using fuel flow sensor and control valve to achieve speed lock.

Optimal mouse sensors andactigraph sensors are combined used to detect the physical stare of the driving person.

III. VEHICLE DETECTION TO SAFE GUARD OWN VEHICLE USING ULTRASONICS

Ultrasonics are fixed all around the car in order to detect the distance of approachingvehicles . The resulting distance is used to calculate the speed of respective vehicles byusing microcontroller. In order to protect the car from beinga deadly crash we settwo crucial zones in the monitoring device. They are Alarming zone and Dangerzone.For every zone whenoutside vehicle is entered alarming tone and danger toneare activated accordingly to alert the driver. Alarming tone alerts the driver when

theapproaching vehicle is 10-15mts far away from our vehicle whereas Danger toneactivates when approaching vehicle comes below 5mts to our car. When the alarmingtone is activated driver will have complete control over his vehicle to safely drive hiscar to a safe place but whereas when danger tone is activated driver needs extrasupport to safe guard himself and his car. So monitoring device automatically getsinformation for a safe place using remaining sensors. As soon as driver founds safeplace he will drive his car to their using Automatic Breaking System (ABS) alongwith involvement of speed lock mechanism.

IV. SPEED LOCK MECHANISM

To limit the speed of a vehicle to certain parameter (say 80kmph), we have to limit the speed of driving shaft i.e.engine (by calculating the angular velocity w using v=rw) without the loss of mechanical efficiency.To achieve that we have to control the fuel flow rate to the engine.

We know that fuel flow is regulated by a valve called butterfly valve which is connected to accelerator.The mechanism used in accelerator is simple constrained link mechanism. So to obtain desired flow rate we simply provide a link mechanism which resists the motion of accelerator pedal further.

V. DROWSINESS DETECTION

The third feature of detecting drowsiness of driver is done by using optical mouse sensors and pulse detection sensor combined. When the driver starts to drive his car the pulse detection sensor plates attached beneath of the steering will automatically detects the pulse state of the driver.

The optical mouse sensor which is arranged in front of the driver will detect the pupil moments and helps to analyzes the drowsiness state of driver. The combined sensors resulting output clarifies whether driver is in perfect condition for driving.If the output clarifies the driver is insleep mode, firstly an alarming tone is played in order to alert the driver. If it fails automatically ultrasonic fetches for a safe place after the fuel supply is completely turned off using an external switch connected to monitoring device in order to bring car to rest position.

VI. MONITORING DEVICE

Monitoring device consist of a microcontroller and a display section. Microcontroller is taken to run three features accordingly. Display screen gives thespeed, distancespecifications of approaching vehicles in vehicle detection case. Alarming speakers are also inserted in monitoring device. Monitoring device plays acrucial role in safe guarding the vehicle as it is the combination of both controllingand display section.

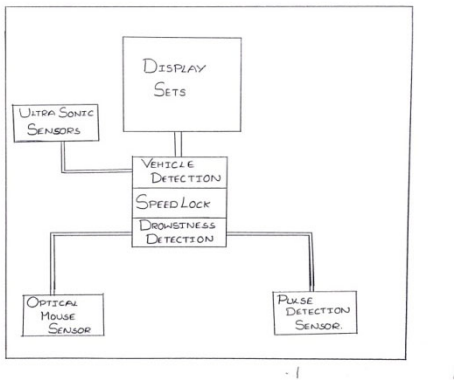


Figure 1: Monitoring device

VII. ALGORITHMS AND EXPERIMENTAL RESULTS

Algorithm for Object Detection

Scan distance from ultrasonic sensor
If dist<=10 or dist>5
{Give alert}
If dist<=5
{ give alert
activate AVM}
else
{Scan distance}

Algorithm for Drowsiness detection

scaneyeBlinkRate from eye blink sensor and pulseRate from pulse sensor
If eyeBlinkRate<12 and pulseRate>=40 or pulseRate<=60{
Give alert sounds
activatespeedlock }

VIII. SIMULATION RESULTS OF DROWSINESS DETECTION

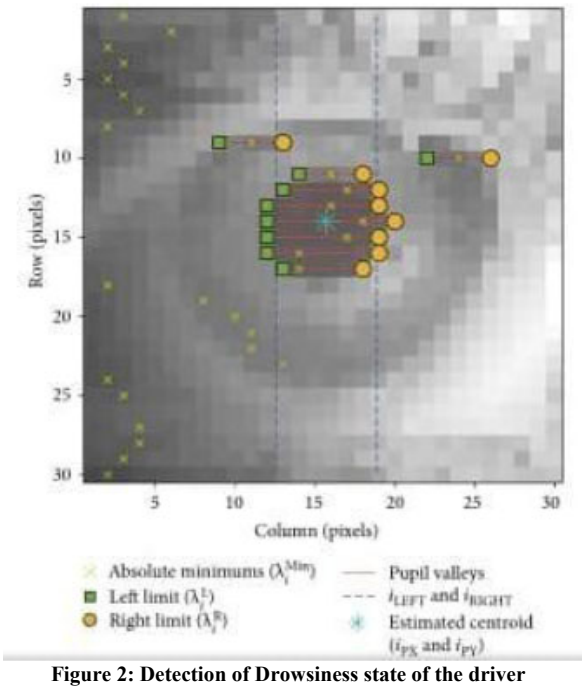


Figure 2: Detection of Drowsiness state of the driver

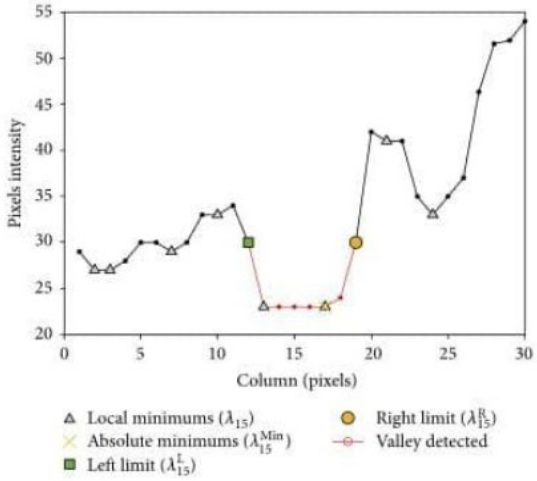


Figure 3: Simulation results of the optical mouse sensor

IX. RESULTS OF VEHICLE DETECTION (i.e.,ultrasonic sensors)

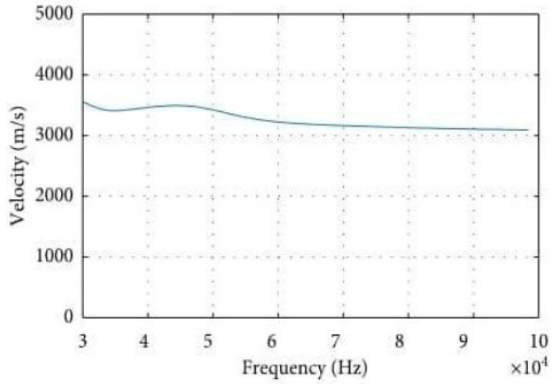


Figure 4.1: Vehicle tracking/detection

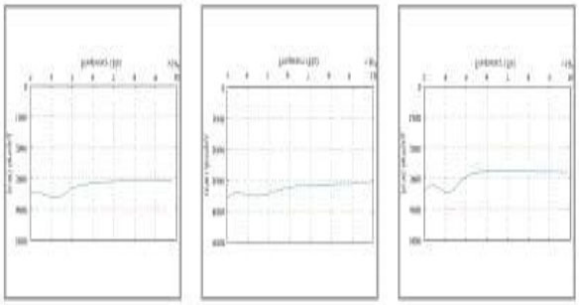


Figure 4.2: Vehicle tracking/detection

X. CONCLUSION

The monitoring electronic device can be embedded in the cars to ensure the safety of the driver and passengers.There are many cases and many possibilities for occurring accidents.Maximum possibilities cover mainly three problems.Using our device we solved these problems like reckless and rash driving,drowsiness,heavy speed. Above mentioned three features are controlled by a microprocessor and monitoring device which always checks the safe place for the vehicle.According to the situations,our monitoring device takes decisions to save us from an accident. If implemented, this device can save many lives in our busy schedule.

REFERENCES

[1] Internal Combustion Engines by V.Ganeshan
[2] https://www.hindawi.com/journals/js/2019/3931713/
[3] https://www.apgsensors.com/about-us/blog/radar-and-ultrasonic-sensors
[4] G.S.Kino(1987) Acoustic Waves in Solid Media,Imaging and Analog Signal Processing,PrenticeHoll,NewJesrsey.

★★★

PREDICTION OF PRODUCT ON THE BASIS OF PRODUCT SELECTED BY CUSTOMER USING APRIORI ALGORITHM

[1]AGAM JAIN, [2]ABHISHEK SABLE

[1,2]Dept. of Computer Science Engineering, Medi-Caps University, Indore, India
E-mail: [1]jaagam520@gmail.com, [2]abhisheksable24@gmail.com

Abstract - Using data mining and association rule mining to find relevant relations among products to increase the sales and generate more revenue is a great practice nowadays. By providing attractive discounts on products being sold together is a masterstroke in increasing sales, which is good and profitable, but there is always room for improvement. There are some products which have high demands in market and the reason behind their frequent sales in not with other products and it is due to their supply and demand. So to provide discounts on such products is, in a way a loss for a retailer. So instead of just following association rules generated by using Apriori algorithm, supply and demand and discount factors can be used to increase sales but moreover increasing the profit margin and will help in arrangement of such products in such a way to make shopping efficient for consumer but also to try giving him a better tour of store and products. This work predicts the product as per the products selected by customer by using the association rules generated using apriori algorithm.

Keywords - Market Basket Analysis, Apriori Algorithm, Discounts, Supply and Demand

I. INTRODUCTION

In today's world of internet, almost every other person is connected to internet or willing to connect. Everyday millions of new users get connected to internet and enjoy the services. With that much number of users, new data is being created every day. This data can be in the form of text, images, videos, audio, graphics, etc and all this data is being shared on internet every day. With a constructive use, this data has become a fuel for many organizations. Nowadays almost every other organization is collecting data of users from many sources and using it according to their need to improve the business of company, providing a better value for its users, etc. To take the full advantage of data the companies need to perform various steps like data preprocessing, data integration, data mining, model building, testing, implementing, etc.

Companies have been collecting data of their customers since long time and therefore they have already collected huge amount of data. Since the users are generating data at much faster pace than a few years back, the collection of data is just keep growing. The more data they have means the better chances they have to understand the behavior of their customers.

Data mining is one of the major steps because this step gives the insight of data. This is the most useful technique helping companies to extract valuable information from huge amount of data. One of the major applications of data mining is in Market basket analysis, where we study patterns between the customers to get the correlation between the different products. Market basket analysis has become very popular in recent times and very helpful as well, for the organizations to understand their customer's behavior and correlation between products purchased in same basket.
Association mining algorithms being used for market basket analysis are Apriori algorithm, Frequent-Pattern growth algorithm, etc. Using such algorithms, relation between products purchases can be found and retailers can take decisions to increase sales and providing more value to consumers.

II. RELATED WORK

Data mining is one of the major processes of finding hidden knowledge and thus finding useful patterns from a huge database. Recently, these practices have attracted many database experts because of its importance in retail industry [1].
One of the very important techniques in data mining is association rule mining which can be really helpful in finding the relationship between data items [2]. Market basket analysis is one of the major application of data mining to analyze items in a basket that customer has for that transaction [3].

Apriori algorithm is being used for market basket analysis and generates item sets and further rules. Concept of mix and match [4] can also be used where related items of different brands can be put together.
So far a lot has been done in research of market basket analysis. Efficient apriori algorithm methods have been introduced with optimizations [5].
Here we can get the required association rules using apriori algorithm but frequent pattern method can also be used [6].

All these studies did not considered supply and demand factor so far to provide discounts. By considering them, the need of offering discounts can be evaluated and unnecessary discounts can be avoided for generating better profits.

III. PROPOSED METHODOLOGY

Keeping in mind the popularity, efficiency and usability of data mining. Concepts of data mining are been used to extract the relevant information from customers data of a company. Huge datasets can be mined using the practices of data mining.

Association rule mining: Relation between various products from customer's basket can be mined and mining of these association rules is the most important step in data mining. For example: Suppose a customer a new 'smart phone' online, now there is s strong probability that the customer will also buy a 'cover' with it. When more frequent buyers will do the same process, Association rule mining will generate a rule by this transaction in the form, "smart phone → cover", where 'smart phone' is antecedent and 'cover' is the consequent. There are measuring terms that plays an important role in this association analysis: Support and Confidence. Both the terms can be defined as –

Support: It is the percentage of transaction that follows the rule i.e. support = (number of transactions containing smart phone and cover together)/(total number of transactions).

Confidence: It is the percentage of transactions containing cover, where smart phone is already there i.e. confidence = (number of transaction containing smart phone and cover together)/(number of transactions containing Smart phone).

Antecedent support: Percentage of transactions containing antecedent of the rule.

Consequent support: Percentage of transactions containing consequent of the rule.

Lift: Measurement of frequency of antecedent and consequent appearing together to the frequency if they were independent.

Conviction: Conviction of rules tells about the effect of antecedent on consequent.

Apriori algorithms: It is a fundamental and important method for finding frequents item sets and association among them. It is basically designed to operate on transactions. It scans the data with repetitive approach of breadth-first-search to generate the frequent item sets. Apriori uses a bottom up approach to generate candidates, and group of candidates are tested against data.

Steps of applied method are discussed below –

1. Load the data into the program.
2. Perform data preprocessing which includes cleaning the data, removing null values, fixing inconsistent data, etc.
3. Encode the transactions in the form required to perform apriori algorithm on them.(hashcode form)
4. Using apriori algorithm calculating minimum support, minimum confidence and minimum lift.
5. Calculate the other matrices of rules like antecedent support, consequent support, lift, conviction, support and confidence for each rule obtained.
6. Calculate dependency factor of rules using lift and conviction.
7. Required rules are ready for various visualizations.

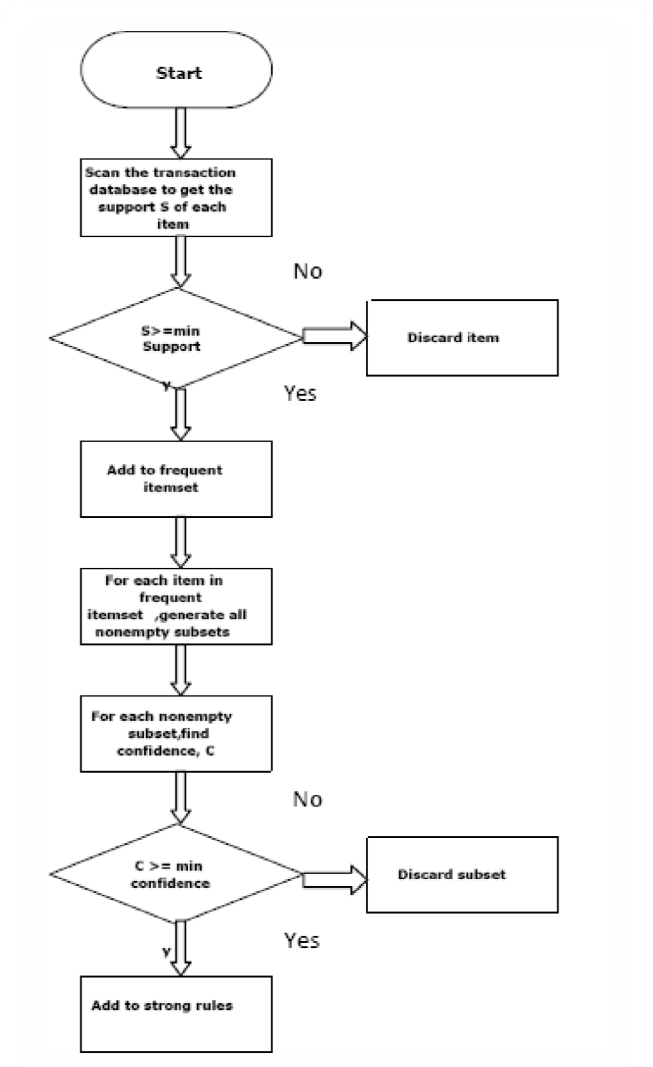


Figure 1: diagram for own steps

IV. USING DISCOUNTS, SUPPLY AND DEMAND

After the rules have been generated there are some rules which will be more important than others by comparing the supply and demand of the items individually. It will also help in ignoring some irrelevant rules that have been generated during association rule mining. Different values of discount can be applied by retailers while implementing the rules and can check the performance of these rules with different discounts. Supply and demand can also be considered while offering discounts, as instead of offering discounts in pair for the products in demand, retailer can offer discount on products of other brands in same category to increase the profit revenue. In

this manner we can use supply and demand to make arrangements of products and predicting a reasonable discount. Discount factor can be used to check the performance of association of rules.

Example: A 'hand sanitizer' is a necessity nowadays and everyone is purchasing it whenever they go for shopping so we can arrange it in shop accordingly and since 'hand sanitizer' of one brand is popular, we can offer discount on product of same category but different brand to increase its sales and we can also apply different discounts to check what discount value can actually improve sales as well as generate good profit.

Efficient use of association rules: In this paper we are proposing how we can efficiently use the obtained association rules by introducing the influence of discount and supply and demand factors. We can study if the particular product from association rule is more popular than others and use it as a medium to give costumer a tour of store or if the product is already in demand than no need to provide discount on it instead retailer can offer discount on same product of different brand with higher profit margin. Performance of association rules can be analyzed on different amount of discount to study the most reasonable discount that can be offered to increase the sales and make more profit margins and keep increasing the sales of products.

Dependency = $\sqrt{(\text{lift}^2 + \text{conviction}^2)}$

V. RESULT

The proposed work has been implemented using python as programming language. The GUI of the program has been shown in Figure 2, which shows its home screen. Here we can input the required fields i.e. minimum support, minimum confidence and minimum lift.

Figure 3 shows the rules and the dependency of consequent on antecedent. Higher value of dependency indicates the better performance of rule. This means the rule is performing well and probability of purchase of consequent is higher after purchase of antecedent. Hence such results can be used to arrange both products in such a way to give customer a short tour of store.

Output on figure 4 displays the required association rules with the values of different matrices. We can examine how in some rules the antecedent is having so much positive effect on the sales of consequent whereas in other rules, it's not the case. In some rules the dependency among two products is higher than others. In graph 1 the matrices of rules can be compared visually how contribution of sales of antecedents and consequents performed individually and in pair as well. High gap between support and confidence bar show the importance of rule. In graph 2 we can visualize the performance of rules and how the sales of consequent are depends on antecedents by considering dependency factor. When sales of one product depend on other and demand of the independent product is higher in that case both products can be kept nearer but no need to offer too much discount because of high dependency, instead discount can be offered on same category product but different brand with better profit margin. If there is no big difference between antecedent support and confidence of rule then the rule holds true and no need to provide too much discount. In the same way the data can be analyzed with different discounts and the performance of rules can be analyzed to get the most profitable discount while keep increasing the sales.

Figure 2: GUI screen

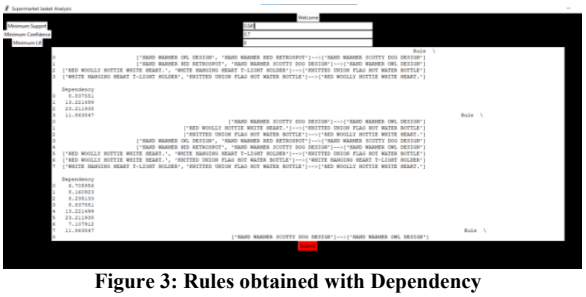


Figure 3: Rules obtained with Dependency

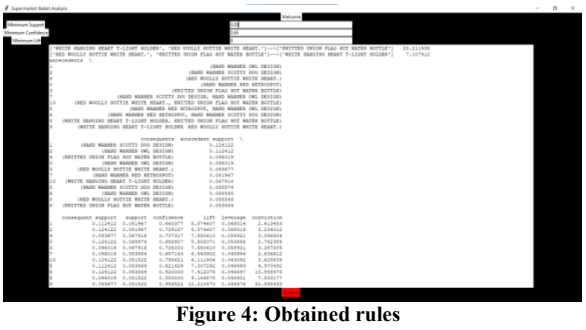
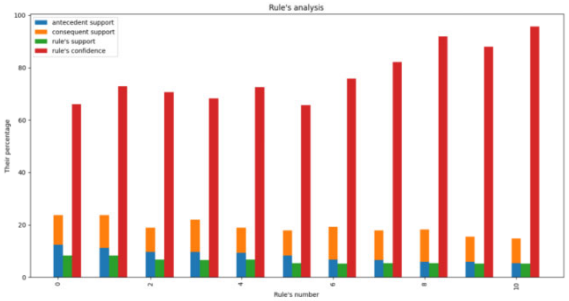
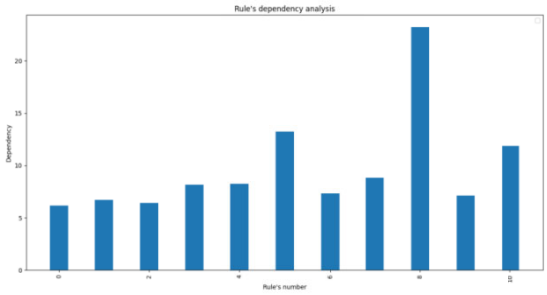


Figure 4: Obtained rules



Graph 1: Rules visualization with different parameters



Graph 2: Rules visualization with Dependency

VI. CONCLUSION

Considering demand and supply and discounts will actually help in long run for all the businesses. .The above discussed method can really be a great approach to generate better offers that will benefit customers as well as the retailers. It will reduce the need of unnecessary discounts and will be helpful in arrangements of products in a store on the basis of supply and demand to make customer spend more time in store exploring wide range of products and shopping more. In this manner businesses can also increase the consumption of products and hence the demand will increase. That will eventually increase the resulting profits.

REFERENCE

[1] Lin, L. and Pei-qi, L. 2001. Study on an Improved Apriori Algorithm and its Application in Supermarket.

[2] S.O. Abdulsalam Department of Computer, Library and Information Science, Kwara State University, Malete, Nigeria : Data Mining in Market Basket Transaction: An Association Rule Mining Approach International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 7– No.10, October 2014

[3] Fachrul Kurniawan, Binti Umayah, Jihad Hammad, Supeno Mardi Susiki Nugroho, Mochammad Hariadi, "Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data", Knowledge Engineering and Data Science (KEDS) pISSN 2597-4602 Vol 1, No 1, January 2018, pp. 20–25

[4] Palash Jain, Naman Kumar Jain, Mohit Premani, Shimpy Goyal, "Super Market Basket Analysis with Mix-&-Match", Proceedings of ieeeforum International Conference, 11th April, 2020, Pune, India

[5] Changxin Song, Research ofAssociation Rule Algorithm based on Data Mining.

[6] Hemant Kumar Soni, Dr. Sanjiv Sharma, Dr. Manisha Jain, Frequent Pattern Generation Algorithms for Association Rule Mining : Strength and Challenges, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016

[7] Dr. M. Dhanabhakyam , Dr. M. Punithavalli, "A Survey on Data Mining Algorithm for Market Basket Analysis". Global Journal of Computer Science and Technology Volume 11 Issue 11 Version 1.0 July 2011

[8] Russell, G. J. and Petersen A. 2000. Analysis of Cross Category Dependence in Market Basket Selection, Journal of Retailing.

[9] Warnia Nengsih, Department of Computer, Politeknik Caltex Riau Riau -Indonesia "A Comparative Study on Market Basket Analysis and Apriori Association Technique", 2015 3rd International Conference on Information and Communication Technology (ICoICT)

★★★

ADVANTAGE OF MAKE-TO-STOCK STRATEGY BASED ON LINEAR MIXED-EFFECT MODEL

[1]YU-PIN LIAO, [2]SHIN-KUAN CHIU

[1]Chairman's Office, Winstar Display Corp., Taichung, Taiwan
[2]Ph.D. Program of Business, Feng Chia University, Taichung, Taiwan ,
[3]Department of International Trade, Feng Chia University, Taichung, Taiwan
E-mail: [1]panjc@winstar.com.tw, [2]skchiu@fcu.edu.tw

Abstract - In the past few decades, demand forecasting becomes relatively difficult because of the rapid changes of world economic environment. In this research, the make-to-stock (MTS) production strategy is applied as an illustration to explain that forecasting plays an essential role in business management. We also suggest that linear mixed-effect (LME) model could be used as a tool for prediction and against environment complexity. Data analysis is based on a real data of order quantity demand from an international display company operating in the industry field, and the company needs accurate demand forecasting before adopting MTS strategy. The forecasting result from LME model is compared to the common used approaches, times series model, exponential smoothing and linear model. The LME model has the smallest average prediction errors. Furthermore, multiple items in the data are regarded as a random effect in the LME model, so that the demands of items can be predicted simultaneously by using one LME model. However, the other approaches need to split the data into different item categories, and predict the item demand by establishing model for each item. This feature also demonstrates the practicability of the LME model in real business operation.

Keywords - Forecasting, Linear Mixed-Effect Model, Make-to-Stock, Order Demand, Production Strategy

I. INTRODUCTION

Demand forecasting is crucial for supply chain management. Production planning, inventory management, and manufacturing scheduling are typically formulated according to short- and long-term expected demand [1].

To reduce the occurrence of delivery delays caused by the "crowding out" effect of manufacturing processes, contemporary enterprises have gradually changed their production patterns from make-to-order (MTO) to make-to-stock (MTS), and increasingly fewer enterprises are using the MTO production strategy [2 , 3]. The MTO production involves commencing product production only after the customer places the order. The MTS production pattern entails a stocking-up production, in which a company manufactures products and stores them in inventory before customer orders are received. Subsequently, the company sells its stock as customer places orders. If a company receives orders requesting a high mix of products but in low volumes, it must be capable of forecasting their order demand accurately before attempting an MTS production strategy.

Accordingly, the advantages of the MTS production strategy—including quick delivery, arranging a long-term manufacturing schedule, reducing the stock levels, and stabilizing product prices—can be realized. Worldwide, variation in customer demand has forced many manufacturers to adopt a high-mix low-volume production model. However, this type of enterprise is not as efficient as a low-mix high-volume enterprise. Therefore, determining how high-mix low-volume enterprises can enhance their business operation performance urgently requires a solution. Hence, accurately forecasting order demand is a fundamental to successfully applying the MTS production strategy to a high-mix low-volume business operation model. Because inaccurate demand forecast is a concern for high-mix low-volume enterprises, the MTO production strategy is typically adopted. However, this production pattern increases financial risks and requires a long delivery time, making centralized production difficult, which subjects production lines to frequent changes, resulting in high operating costs and low product quality. Complex operations are the primary cause of human error and low job satisfaction.

Therefore, if the inefficiency of the high-mix low-volume business operation model cannot be solved, then, despite a high business revenue, business operation costs would increase rapidly, product quality would reduce, and employee job satisfaction and customer satisfaction would decrease, which result in that business development would stagnate. Therefore, the forecasting method proposed in this study can provide a crucial basis for transitioning from using the MTO to the MTS production, and may offer a viable solution for improving the business operation performance of high-mix low-volume enterprises.

The application and improvement of the proposed forecasting method can assist researchers with understanding the characteristics of business operations and construct related business operation models. Forecasting ability depends on crucial information and reliable forecasting methods. In recent years, demand forecasting has become increasingly complex, primarily because the global

economic environment has gradually changed. The underlying reasons for this change can be explained in terms of the following four dimensions: volatility, uncertainty, complexity, and ambiguity (VUCA) [4 , 5 , 6], all of which have been shown to influence demand forecasting [7]. Volatility means that new products are rapidly developed, product lifecycles are shortened, customer preferences change suddenly, and organizations are frequently restructured; consequently, historical data diminishes in value. Uncertainty refers to unknown factors that cause sudden shifts in demand, and these factors are generally regarded as outliers or interferences. Complexity means that the interaction of these influential factors cannot be modelled easily, and ambiguity refers to fuzzy events and situations that cannot be quantifiably defined, leading to the loss of key influential factors. In summary, according to the influence of economics on demand forecasting, developing a reliable forecasting method requires analyzing whether historical data can contribute to demand forecasting, and whether the effects of influential factors can be identified. To meet the requirements of modern forecasting methodologies, this study proposed using linear mixed-effect models to perform forecasting. Linear mixed-effect models have been extensively developed and widely applied in various fields. However, no study has used this model to forecasting in business operation. Linear mixed-effect models are characterized by the inclusion of temporal factors and explanatory variables and the analysis of their significance. Accordingly, crucial influential factors can be identified to forecast demand. These characteristics fulfill the requirements of modern forecasting methodologies and can be used as the basis for companies to improve their operation efficiency and to develop competitive advantages.

The following sections explore the influences of the MTO and MTS production strategies on business operation as well as the role of forecasting in the MTS strategy, provides a review of the literature on forecasting methodologies, and summarizes the strengths and weaknesses of commonly used forecasting methods. In addition, the proposed linear mixed-effect model as well as a method for model parameter estimation are introduced. Subsequently, the order demand of a manufacturer in central Taiwan is forecasted using product type as a crucial explanatory variable. Specifically, the linear mixed-effect model is applied to forecast the order demand for 20 individual product types. A 1-year forecast of monthly demand is reported, and three types of forecast errors are used to assess the forecasting ability of the model. The results show that the forecasting ability of the linear mixed-effect model in an empirical analysis is superior to those of a linear forecasting model, exponential smoothing method, and time-series forecasting method.

II. LITERATURE REVIEW

A. Influences of the MTO and MTS on Business Operations

Modern production strategies primarily involve two main production patterns: the MTO (based on customer orders), and the MTS (based on production capacity) [8]. From the perspective of customers, one competitive advantage of the MTS production is short delivery time and quick response [9]. Therefore, identifying the types of products that are specifically suitable for the MTS production pattern or both MTS and MTO patterns is a favored research topic in management science [8].

Regarding the influences of the MTO and MTS production strategies on business operations, Hendry and Kingsman [10] showed that the MTS and MTO production strategies are mostly used for manufacturing standard and customized products, respectively. Regarding the attributes of orders, order demand for MTS products is generally predictable, whereas that for MTO products is irregular and unpredictable. Concerning production planning, MTS production lines operate according to forecast results, and the production line schedule can be adjusted easily. However, the schedule of MTO production lines is determined based on recent order demand, and long-term manufacturing schedules are difficult to determine. In terms of product delivery, enterprises that adopt the MTS production strategy can ensure rapid product delivery, thus maintaining high customer satisfaction. The MTO production pattern requires long delivery times, and enterprises adopting this strategy must communicate with customers to achieve consensus regarding product delivery time. Concerning product price, compared with prices of products produced adopting the MTO strategy, the prices of MTS-produced products are relatively more stable. Soman, van Donk, and Gaalman [8] indicated that the MTO production pattern is effective for handling orders requesting high-mix customized products; the production planning for the MTO strategy must prioritize meeting order demands, while production effectiveness is determined according to crucial elements in the orders (e.g., the expected delivery volume and number of delayed delivery days). The goal of a company that manufactures MTO products is to shorten product delivery times; production efficiency emphasizes the importance of capability planning, orders that are lost due to problems with manufacturing processes, and on-time product delivery. By contrast, the MTS production pattern is effective for handling uniform product specifications and less customized products, where production planning is determined based on product demand forecasting and production effectiveness is production-oriented. Therefore, the goal of a company manufacturing MTS products is to enhance product availability, and its production efficiency emphasizes the importance of inventory policy,

finished goods inventory, one-off or batch production, and accurate demand forecast. Rajagopalan [11] indicated that inventory costs are slightly higher for the MTS strategy than for the MTO strategy, particularly for one-off and batch production.

In summary, the MTS strategy relies heavily on the accuracy of product demand forecasting. Because of accurate forecasting, the advantages of the MTS production strategy, including short delivery time, manageable long-term manufacturing schedule, and stable product prices, can be realized. In addition, accurate forecasting can optimize inventory levels; therefore, companies applying the MTS strategy can effectively control inventory costs. Some researchers have explored the inventory policies and material control mechanisms in MTO production [12]. The forecasting method proposed in this study provides a relatively accurate basis for forecasting random customer orders (demand) for MTS production.

B. Forecasting Methodology

Two main types of forecasting methodology exist: (1) statistical methods; and (2) data mining and machine learning [13]. Both types of forecasting methodology are aimed at identifying the relationship between influential factors (independent variables) and research variables (dependent variables), and identifying the effects of the influential factors on research variables [7]. These two methodologies involve distinct approaches to interpreting analysis models. The statistical methodology is based on the data derived from a specific mathematical model as well as unobservable errors. The machine-learning methodology avoids fitting data to a specific model and develops algorithms that are suitable for various types of data. These two methodologies differ in their strengths and characteristics [13]. The statistical methodology uses the probability distribution of errors to infer the significance of the influential factors in a model. The reliability of inferences correlates positively with the mathematical model. The machine learning methodology uses the size of forecast errors as a basis for selecting the optimal forecasting model.

Several typical forecasting methods are introduced as follows, the characteristics of which are shown in Table 1. The exponential smoothing method was proposed by Holt [14] and the statistical theoretical foundation for this method was established by Muth [15]. This method involves using a demand observation and predictive value in the current period to determine the predictive value for the subsequent period by using weighted mean. To date, the exponential smoothing method has been widely applied to forecast demand under the bullwhip effect [16] and to plan inventory control strategies [17]. Moreover, the methodology for exponential smoothing has been developed in recent years into one that incorporates the effect of influential factors on the accuracy of demand forecasts [7, 18, 19]. Wang [19] used a model selection method where crucial influential factors were included in the selected model, and nonsignificant factors were removed to avoid over-fitting the model.

Time-series model was first developed in the nineteenth century, and past studies related to such model were then systematically compiled by Box and Jenkins [20] into a book. A time-series autoregressive integrated moving average (ARIMA) model integrates an autoregressive process and moving average process after obtaining a finite difference from time-series data. The ARIMA model is used to estimate the correlations parameter between the time points of observed values, and the estimated parameter values can then be used for forecasting. Subsequently, Box and Tiao [21] added other time-series influential factor to the ARIMA model. Pankratz [22] called this model the dynamic regression model.

Forecasting method	Can handle temporal data	Can include influential factors	Analyzing the importance of influential factors (e.g., *p* value)
Linear mixed-effect model	○	○	○
Exponential smoothing method	○	△	△
ARMA	○	△	△
Linear model	△	○	○

Table 1. CHARACTERISTICS OF F ORECASTING M ETHODS . (○: YES ; ▵: YES FOLLOWING MODIFICATION BY OTHER STUDIES

Linear regression models are a type of linear model that are most frequently mentioned in statistical analyses. Linear models assume that research variables and influential factors are linearly related, and thus can be used to explore the effect of influential factors on research variables. Furthermore, linear models assume that observation values are mutually independent; thus, this model is applicable for analyzing data containing mutually independent observation values. If linear models are used to analyze time-correlated data, i.e., the observation values being correlated over time, then unbiased but invalid model coefficient estimators can be obtained. Consequently, the standard errors of the model coefficient estimators would be incorrect, and problems regarding statistical testing within the models arise, such as whether the model coefficients are significantly greater than 0, whether the models exhibit explanatory power, and whether the predictive intervals are reliable in forecast analysis [23, 24].

Linear mixed-effect models can be considered as an extension of linear models. The linear mixed-effect

models add random effects to linear models with fixed effects. Hence, a model that has both fixed and random effects is called a linear mixed-effect model. Linear mixed-effect models are typically used to describe the relationship between research variables and categorical factors with correlated observation values. A characteristic of the mixed-effect models is that observation values at the same categorical level have identical random effect values for dependent variables; observation values at different levels have distinct values of random effect. This characteristic explains the correlation between observation values at an identical level. Therefore, linear mixed-effect models differ considerably from linear models. The mixed-effect model can be applied to data where observation values are correlated (e.g., longitudinal data, repeated measures data, and multilevel data). However, linear models can be applied only to data where the observation values are mutually independent. In industrial operations, the pattern of data observations is often time-correlated. For example, when forecasting monthly product demand or monthly inventory levels, the observation values are correlated over time. Under such circumstances, the linear mixed-effect model is more accurate than linear models for identifying statistically significant factors.

In the past 2 years, the linear mixed-effect model has been broadly applied in various fields, such as the timber industry [25], medicine [26, 27], and ecology [28], to identify crucial influential factors. In addition, numerous studies have established models for forecasting [29, 30]. However, in industrial engineering and management science [24, 31, 32, 33], no study has used the linear mixed-effect model to make predictions by using time-correlated data or to identify key influential factors. Therefore, in this study, a linear mixed-effect model was applied to business operations to analyze the importance of influential factors, and to forecast product demand; in addition, the performance of the linear mixed-effect model was compared with that of other methods, which are the research contributions of this study.

III. LINEAR MIXED-EFFECT MODEL

According to parameter attributes, two types of effect exist in a linear mixed-effect model: fixed and random effects [34, 35]. In a linear model, the parameters are all fixed values and therefore its corresponding covariates are referred to as fixed-effect parameters. The fixed effect describes the true value of the coefficient for an entire population, or the true value of the coefficient for a factor that can be repeatedly tested under identical conditions. If a factor in a model exhibits a random effect, then the factor is sampled from an entire population. The random effect is a coefficient of the factor; moreover, the coefficient is a random variable and not a fixed value. The following section introduces the linear mixed-effect model developed by Laird and Ware [36] and the estimation of model parameters, and describes how the research variables are forecasted.

A. Linear Mixed-Effect Model

In contrast to a multilevel model, a single-level linear mixed-effect model [36] was employed in this study. The multilevel model differs from the single-level model in terms of the covariance matrix of the observation values. The single-level model involves only one level, whereas the multilevel model involves at least two levels. The covariance matrix of the multilevel model is more complex than that of the single-level model. In practice, whether using a single-level or multilevel model is more appropriate depends on the data structure of the observation values. Although the covariance matrices of the two models differ, the observation values of the various groups at a fixed level are independent of each other, and the within-group observation values are intercorrelated. In the multilevel model, a group at one hierarchy level becomes the next level of the hierarchy.

The single-level linear mixed-effect model developed by Laird and Ware [36] is expressed as follows:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, M \tag{1}$$

$$\mathbf{b}_i \square N(\mathbf{0}, \boldsymbol{\Psi}), \boldsymbol{\varepsilon}_i \square N(\mathbf{0}, \boldsymbol{\Lambda}_i), \tag{2}$$

where $\mathbf{b}_i$ is a matrix that is independent of $\boldsymbol{\varepsilon}_i$ (index i denotes the ith group at a single level), $\mathbf{y}_i$ contains n_i observation values for the ith group, M denotes the number of groups, $\boldsymbol{\beta}$ denotes a p-dimensional vector for the fixed effect, $\mathbf{b}_i$ denotes a q-dimensional vector for the random effect, $\mathbf{X}_i$ denotes an $n_i \times p$ covariance matrix for the fixed effect, $\mathbf{Z}_i$ is an $n_i \times q$ covariance matrix for the random effect, and $\boldsymbol{\varepsilon}_i$ denotes an n_i-dimensional within-group random error term. The variable $\boldsymbol{\varepsilon}_i$ obeys a multivariate normal distribution with an expected value of 0 and a covariance matrix of $\boldsymbol{\Lambda}_i$, and $\mathbf{b}_i$ obeys a multivariate normal distribution with an expected value of 0 and a covariance matrix of $\boldsymbol{\Psi}$. The model assumes that $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_j$ are mutually independent ($i \neq j$); in addition, $\boldsymbol{\varepsilon}_i$ and $\mathbf{b}_i$ are mutually independent. Therefore, considering Models (1) and (2), the covariance matrix of the within-group observation values $\mathbf{y}_i$ is expressed as follows:

$$\mathbf{V}_i \equiv \mathrm{Var}(\mathbf{y}_i) = \mathrm{Var}(\mathbf{Z}_i\mathbf{b}_i) + \mathrm{Var}(\boldsymbol{\varepsilon}_i) = \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \boldsymbol{\Lambda}_i \tag{1}$$

where the nondiagonal elements of $\mathbf{V}_i$ are not required to be 0. Therefore, according to (3), Models (1) and (2) allow the existence of the correlation between observation values within a group. This is a major difference that the two models have with the

linear model.

B. Estimation of the Model Parameters

This section introduces estimation methods that adopt the linear mixed-effect model: the maximum likelihood (ML) and restricted ML (REML) estimation methods. Regarding the ML method, the estimates of ML estimators are those that reach the maximum value of ML functions. By comparison, the REML method is aimed at identifying the estimators that exhibit unbiased characteristics. Therefore, estimators obtained using the REML method are unbiased, whereas those derived using the ML method could feature either biased or unbiased property. Therefore, most researchers prefer the REML method [34, 35]. We introduce the estimation procedures for both of these estimation methods, although only the REML method was used in this study.

First, the model $\boldsymbol{\beta}$ coefficient and covariance matrix of observation values $\mathbf{V}_i$ are estimated as follows. In Models (1) and (2), the expected values of $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be 0; thus, the expected value of y_i is $X_i\beta$ (i.e., $E(\mathbf{y}_i)=\mathbf{X}_i\boldsymbol{\beta}$). Because the covariance matrix of y_i is V_i (i.e., $Var(\mathbf{y}_i)=\mathbf{V}_i$) and because $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ obey an independent multivariate normal distribution, the marginal distribution of y_i is a multivariate normal distribution expressed as follows:

$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta},\ \mathbf{V}_i)$$

The ML function is expressed as follows:

$$L(\boldsymbol{\beta},\boldsymbol{\theta})=\prod_{i=1}^{M}(2\pi)^{\frac{-n_i}{2}}\det\left(\mathbf{V}_i\right)^{\frac{-1}{2}} \times\exp\left\{-\frac{1}{2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right\}$$

where $\boldsymbol{\theta}$ denotes the set of $\mathbf{V}_1,...,\mathbf{V}_M$. To facilitate differentiation, the natural logarithm of the ML function is used instead of the ML function to evaluate the ML and REML estimators, and define $l(\boldsymbol{\beta},\boldsymbol{\theta})=\ln L(\boldsymbol{\beta},\boldsymbol{\theta})$. ML estimation method The ML estimates of β and θ are the values that maximize $l(\boldsymbol{\beta},\boldsymbol{\theta})$ and thus are also the values that maximize L(β,θ). Calculating the maximum value of $l(\boldsymbol{\beta},\boldsymbol{\theta})$ is challenging. Typically, let $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, and evaluate the value of β such that it maximizes $l_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}(\boldsymbol{\beta},\boldsymbol{\theta})$. Subsequently, let $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}$, and calculate the value of θ such that it maximizes the value of $l_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta},\boldsymbol{\theta})$. This process is iterated until the change in $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ is within a tolerance error (i.e., the $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ values converge).

Specifically, we first let $\boldsymbol{\theta}$ be $\hat{\boldsymbol{\theta}}$ (equivalent to letting $\mathbf{V}_i$ be $\hat{\mathbf{V}}_i$, $i=1,...,M$). Under these conditions, $\mathbf{y}_i$ obeys $N(\mathbf{X}_i\boldsymbol{\beta},\hat{\mathbf{V}}_i)$. An analytical solution for $\boldsymbol{\beta}$ can be obtained by using the generalized least squares method.

$$\hat{\boldsymbol{\beta}}=(\sum_i\mathbf{X}_i^T\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1}\sum_i\mathbf{X}_i^T\hat{\mathbf{V}}_i^{-1}\mathbf{y}_i \qquad (4)$$

Accordingly, $l_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}(\hat{\boldsymbol{\beta}},\boldsymbol{\theta})$ is the maximum value. Next, fix $\boldsymbol{\beta}$ in $l(\boldsymbol{\beta},\boldsymbol{\theta})$ as $\hat{\boldsymbol{\beta}}$, denoted by $l_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta},\boldsymbol{\theta})$, to obtain a $\boldsymbol{\theta}$ that maximizes the value of $l_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta},\boldsymbol{\theta})$, where

$$l_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta},\boldsymbol{\theta})=-\frac{1}{2}\left(\sum_i n_i\times\ln(2\pi)+\sum_i\ln(\det(\mathbf{V}_i))+\sum_i(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})^T\mathbf{V}_i(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})\right) \qquad (5)$$

where $\mathbf{V}_1,...,\mathbf{V}_M$ are functions of $\boldsymbol{\theta}$. Typically, $l_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta},\boldsymbol{\theta})$ is not a linear function for $\boldsymbol{\theta}$. Consequently, no analytical solution for $\boldsymbol{\theta}$ exists, and an algorithm must therefore be used to obtain a numerical solution for $\boldsymbol{\theta}$. Commonly used algorithms include the expectation-maximization (EM) algorithm, Newton's method, and Fisher's scoring algorithm. Previous studies have described these algorithms in detail [36, 37, 38], including a comparison of their strengths and weaknesses [35]. An algorithm can be used to obtain a numerical solution for $\boldsymbol{\theta}$ (i.e., $\hat{\boldsymbol{\theta}}$), the result of which can be converted to $\hat{\mathbf{V}}_i$. Subsequently, the calculation is performed iteratively by using Equations (4) and (5) until the values of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ converge.

REML estimation method The REML method is another approach for estimating $\boldsymbol{\theta}$. The REML estimate of $\boldsymbol{\theta}$ is obtained by applying an iterative method to a restricted natural-logarithm ML function.

$$l_{REML}(\boldsymbol{\theta})=-\frac{1}{2}\left((\sum_i n_i-p)\times\ln(2\pi)+\sum_i\ln(\det(\mathbf{V}_i))+\sum_i(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})^T\mathbf{V}_i(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})+\sum_i\ln(\det(\mathbf{X}_i^T\mathbf{V}_i\mathbf{X}_i))\right) \qquad (6)$$

Regarding the difference between the restricted natural-logarithm ML function (6) and Equation (5), Equation (6) accounts for the loss in degrees of freedom. Therefore, the estimator of $\boldsymbol{\theta}$ obtained using the REML is an unbiased estimator. The REML method involves applying Equation (4) to obtain the

estimator of $\boldsymbol{\beta}$. For the REML, Equations (4) and (6) are iteratively used until the values of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ converge. Equation (4) is used in both the ML and REML estimation methods to estimate $\boldsymbol{\beta}$. However, the functions employed to estimate θ (i.e., the ML and REML methods use Functions (4) and (6) to estimate $\boldsymbol{\theta}$, respectively) differ between these methods, and they thus yield different values for $\hat{\boldsymbol{\theta}}$. In addition, because $\hat{\mathbf{V}}_i$ is a function of $\hat{\boldsymbol{\theta}}$, different values are obtained for $\hat{\mathbf{V}}_i$; consequently, different $\hat{\boldsymbol{\beta}}$ values are obtained through using these two methods.

Estimating random effect parameters Given $\mathbf{b}_i$, the following equation can be derived from (1):

$$\mathbf{y}_i \mid \mathbf{b}_i \overset{d}{=} N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Lambda}_i)$$

where "$\overset{d}{=}$" represents "distribution equals" and $\boldsymbol{\Lambda}_i$ is given by (2). Therefore, the generalized least squares method can be applied to estimate $\mathbf{b}_i$, which is equal to $(\sum_i \mathbf{Z}_i^T\boldsymbol{\Lambda}_i^{-1}\mathbf{Z}_i)^{-1}\sum_i \mathbf{Z}_i^T\boldsymbol{\Lambda}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$. In the equation, $\boldsymbol{\Lambda}_i$ (a function of $\boldsymbol{\theta}$) and $\boldsymbol{\beta}$ are true values. Therefore, by substituting the ML or REML estimates (i.e., $\hat{\boldsymbol{\beta}}$ or $\hat{\boldsymbol{\Lambda}}_i$), we can obtain the estimator of $\mathbf{b}_i$ as follows:

$$\hat{\mathbf{b}}_i = (\sum_i \mathbf{Z}_i^T\hat{\boldsymbol{\Lambda}}_i^{-1}\mathbf{Z}_i)^{-1}\sum_i \mathbf{Z}_i^T\hat{\boldsymbol{\Lambda}}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}).$$

C. Forecasting Research Variables

After the explanatory variables $\mathbf{X}_i^{new}$ and $\mathbf{Z}_i^{new}$ have been obtained, the estimates of $\boldsymbol{\beta}$ and $\mathbf{b}_i$ (i.e., $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}_i$) described in the previous section can be used to forecast the research variable $\mathbf{y}_i$. The predictive value is as follows:

$$\hat{\mathbf{y}}_i = \mathbf{X}_i^{pred}\hat{\boldsymbol{\beta}} + \mathbf{Z}_i^{pred}\hat{\mathbf{b}}_i. \quad (7)$$

IV. A CASE STUDY

This study adopted a single-level linear mixed-effect model to forecast product demand. In the case study, the sample was a leading professional industrial LCD/OLED display manufacturer. This manufacturer produces products that are critical components of various devices used in daily life and are applied in various industries. Moreover, the company has an international customer base. Table 2 shows the number of orders, total product demand, average product demand per order, and quantity of finished goods from 2009 to 2013. Before 2013, the manufacturer produced more than 5,000 product types, and the average quantity of products required in an order was approximately 400. Thus, the manufacturer is considered to be a suitable example of a business that produces a diverse combination of high-mix products.

A characteristic of high-mix low-volume manufacturers is that they typically commence production only after receiving a customer order. This production pattern is typical of the MTO production pattern, which is mainly adopted to serve customers in niche markets. In recent years, the manufacturer's profits have decreased despite an increasing revenue and market share. Therefore, the manufacturer aimed at changing its production strategy by adopting the MTS production strategy for some product types in order to increase its batch production capacity, reduce its production costs, and improve its production efficiency. In addition, the manufacturer believed that adopting the MTS production strategy would enhance customer satisfaction by ensuring the rapid delivery of customer orders, thereby providing a competitive advantage. Thus, being able to accurately forecast product demand was crucial. Following evaluation, to test the implementation of the MTS production strategy, this study selected the top 20 standard finished products that were most frequently ordered between 2011 and 2013 by customers of the sample manufacturer. As shown in Figure 1, these 20 standard products accounted for 20% of the manufacturer turnover for standard products in 2013, with 86 orders placed in the same year. After implementing the MTS production strategy, the manufacturer planned to run production of each product type once per month per year. Accordingly, the production frequency, cost of handling orders, and frequency of changing production lines was reduced. Thus, its long-term production capacity plans can be implemented to maximize the benefits of producing a high volume of products with fewer runs.

Year	Number of orders	Total product demand	Average product demand for an order	Finished goods quantity
2009	12,929	3,603,141	278.69	2,727
2010	17,968	8,343,884	464.37	3,518
2011	20,169	6,721,194	333.24	4,546
2012	22,589	8,062,890	356.94	5,822
2013	22,361	9,045,056	404.50	5,468

Table 2. NUMBER OF ORDERS AND PRODUCT DEMAND

A. Data Structure

The data structure comprised 20 types of standard finished products. The monthly product demand data were collected from January 2007 to December 2013 for each product type (see S1 Table). The historical

data before 2012 were used to estimate model parameters, and the model was used to forecast the product demand for 2013 (January–December). Not all 20 products were manufactured from 2007. The historical data used to estimate model parameters comprised 1295 observation values (64 observation values on average for each product type). The product lifecycle varied by year, and the product demand varied by month. Therefore, year and month were crucial predictors. For each type of product, the monthly product demands in each month were related. In this study, the explanatory variables (year and month) were added to the linear mixed-effect model to analyze the monthly product demand data. Regarding product sales, the product demand varied by product type. Accordingly, product type was regarded as a crucial categorical variable because of its influence in forecasting the product demand. In this study, according to the characteristics of the mixed-effect model, we used product type as a random-effect term and included the demand for each product type in a universal model to

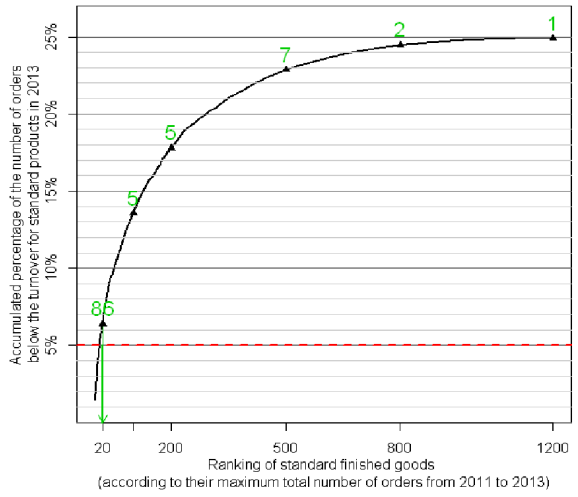


Figure 1. Maximum total number of orders (2011–2013). The plot shows that the accumulated percentage of the maximum total number of orders from 2011 to 2013 is less than the turnover of standard products in 2013. The first 20 products accounted for approximately 20% of the turnover for of standard products. The numbers in green denote the number of orders for standard products in 2013 corresponding to the horizontal axis. forecast the demand for type separately. Subsequently, we compared other commonly used forecasting methods. Unlike the mixed-effect model, other methods did not have a universal model to account for 20 unique product types. Therefore, for the other forecasting methods, the data are required to be divided into multiple data sets according to product type, and the partitioned data are then applied to the forecasting methods depending on the product type for analysis and forecasting. This approach substantially reduces the sample size, reducing the accuracy of the forecast.

B. Model Development

Product demand differed by product type, and thus we assumed the demand for each type of product to be mutually independent. In Model (1), which is the single-level model, random effect was set to be product type, thus yielding various random-effect coefficient for each product type. The model is expressed as follows:

$$\begin{aligned}\mathbf{y}_i &= \beta_0 + \beta_1 \times (\text{year-2007}) \\ &+ \beta_2 \times (\text{year-2007})^2 + \text{month} \times \boldsymbol{\beta}_3 \\ &+ b_{i0} + b_{i1} \times (\text{year-2007})^2 + \boldsymbol{\varepsilon}_i\end{aligned} \quad (8)$$

where $\mathbf{y}_i$ is a vector that denotes the monthly product demand (the vector length is equal to the data quantity for product i); β_0, β_1, β_2, and β_3 denote the intercept, year, year-squared, and month for the fixed-effect term; and b_{i0} and b_{i1} denote the intercept and year-squared for the random-effect term. In Model (8), year was considered as a continuous variable with 2007 used as the baseline. Month was a categorical variable; therefore, the month term in Model (8) was a dummy variable. The dummy variable for month had 11 indicator variables with a value of 0 or 1, and the total product demand in January was used as the baseline. Expressing Equation (1) as Model (8), the fixed-effect explanatory variable $\mathbf{X}_i$ is a matrix comprising a column of 1's vector for the intercept, year, year-squared, and month covariates. Thus, the expression $\boldsymbol{\beta}=[\beta_0\ \beta_1\ \beta_2\ \boldsymbol{\beta}_3^T]^T$ is a 14 × 1 vector, where $\boldsymbol{\beta}_3$ is the coefficient of the dummy variable for the month covariate and has 11 elements. To account for the various product types, we chose the intercept and year-squared covariate as the random-effect explanatory variable, where the intercept was used to account for the average difference of demands between product types, and the year-squared covariate was used to consider the difference between product demands decreased or increased over time. The explanatory variable $\mathbf{Z}_i$ in the random-effect explanatory variable comprised the intercept and year-squared covariate, of which the coefficients are a 2 × 1 vector expressed as $\mathbf{b}_i=[b_{i0}\ b_{i1}]^T$. In Model (8), the year-squared covariate in the random-effect explanatory variable was also a part of the fixed-effect explanatory variable, and was used to account for the fact that the expectation of $\mathbf{b}_i$ was probably unequal to 0; thus, the assumption that $\mathbf{b}_i$ in (2) was equal to 0 was reasonable. The year-squared covariate was included to prevent the annual growth trend from being linear, which enabled the model to more accurately reflect the current situation. The year-squared covariate is crucial to practical operations. The year and year-squared covariates added into the fixed-effect explanatory variable facilitated

establishing a grand model for the 20 product types. The year and year-squared covariates for the fixed effect indicated the average growth trend for the 20 product types, whereas the random effect reflected the specific annual growth trends for each product type. To forecast the monthly product demand for 2013, 2013 was used as the value for the year and year-squared covariates. Both covariates and the target month were input into the explanatory variable to form $\mathbf{X}_i^{new}$ and $\mathbf{Z}_i^{new}$. Subsequently, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}_i$ in (7) were used to obtain the forecasted value $\hat{\mathbf{y}}_i$.

C. Other Forecasting Methods

Comparing forecasting methods is crucial in methodological studies [39 , 40 , 41 , 42 , 43]. The model proposed in this study was compared with commonly used statistical forecasting methods, beginning with the following linear model:

$$Y_j = \alpha_0 + \alpha_1 \times (\text{year}_j\text{-}2007) + \alpha_2 \times (\text{year}_j\text{-}2007)^2 + \alpha_3 \times \text{month}_j + \delta_j \quad (9)$$

where α_0, α_1, α_2, and α_3 are regression coefficients and α_3 denotes the coefficient of the dummy variable for the month covariate, and δ_j is the error term. Model (9) (i.e., the linear model) includes only the fixed-effect term in Model (8) (i.e., the mixed-effect model); therefore, Model (9) was compared with Model (8) to examine the differences when the random-effect term is present or absent in the model. A total of 1295 observations of monthly product demand ($Y_j, j = 1,...,1295$) were used to estimate the coefficients in Model (9) and the significance of the coefficients with P values. In the Results section, Models (8) and (9) are compared regarding forecast accuracy and the P values.

Explanatory variable	Linear mixed-effect model			Linear model		
	Coefficient	Standard error	P value	Coefficient	Standard error	P value
The intercept term	39.46	320.14	.9019	174.54	294.07	.5529
(Year-2007)	800.55	133.52	.0000 ***	746.35	153.69	.0000 ***
(Year-2007)^2	-99.97	25.82	.0001 ***	-93.23	27.62	.0008 ***
February	206.41	283.03	.4660	171.14	327.19	.6010
March	736.88	281.57	.0090 **	716.23	325.51	.0280 *
April	753.30	281.56	.0076 **	762.56	325.50	.0193 *
May	536.51	280.89	.0564 -	504.53	324.70	.1205
June	253.56	281.62	.3681	218.06	325.53	.5031
July	591.73	271.46	.0295 *	556.53	313.77	.0764 -
August	91.35	271.48	.7366	56.21	313.77	.8579
September	711.75	271.46	.0088 **	664.40	313.77	.0344 *
October	297.69	271.05	.2723	255.00	313.28	.4158
November	473.91	272.52	.0823 -	432.47	314.94	.1699
December	360.30	270.62	.1833	308.34	312.71	.3243

Table 3. Linear Mixed-Effect Model Versus the Linear Model. " - ": p < .1; " * ": p < .05; " ** ": p < .01; " * ": p < .001.**

Next, the model proposed in this study was compared with the exponential smoothing method, in which the product demand observation values Y_t's and its predictive values F_t's were used to obtain the predictive values for the subsequent period by calculating a weighted mean. The forecast formula is as follows:

$$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t$$

where α is the weighted coefficient. To accurately forecast the monthly product demand in this case, we adjusted the exponential smoothing method to account for two influential factors (i.e., month and product type). The data were divided into 20 data sets according to each product type, and each data set was divided into 12 subsets (one for each month). For each product type, no more than six observations from each month in the historical data were used. The pre-2012 monthly product demand data were used to forecast the product demand for the corresponding months in 2013. The weighed coefficient was $\alpha = \frac{1}{2(N+1)}$, where N is the number of observations for a month ($N \le 6$).

Finally, the model proposed in this study was compared with a seasonal time-series model; specifically, the autoregressive moving average model

$(ARMA(2,2)_{12})$, which was considered to be a suitable model because the data were not nonstationary time-series data. The mathematical model for ARMA $(p,q)_s$ is expressed as follows:

$$(1 - \sum_{i=1}^{p} \phi_i B^{s \times i}) Y_t = (1 + \sum_{i=1}^{q} \theta_i B^{s \times i}) \xi_t$$

where ϕ_i is the ith order autoregressive process coefficient, B is a backward shift operator, θ_i is the ith order moving-average process coefficient, ξ_t is a normally distributed confounding term, and s is a seasonal parameter. Longitudinal data were collected for each of the 20 product types. A time-series model was established for each of the 20 product types. In this case, the month was regarded as a crucial influential factor for forecasting and thus the seasonal parameter s was set to 12, which indicates the existence of correlations in the data for every 12 month. The samples were categorized by product type, yielding an average of 64 samples for each type of product. The parameters p and q were determined based on the characteristics of an autocorrelation function, a partial autocorrelation function, and an extended autocorrelation function (p = 2 and q = 2). Finally, the $ARMA(2,2)_{12}$ model was used to forecast the product demand for each product type.

	MAE		MAPE		RMSE	
	M	*SD*	*M*	*SD*	*M*	*SD*
Linear mixed-effect model	1,412.71	1,500.04	1.52%	1.50%	1,849.42	1,919.86
Linear model	1,828.96	2,091.93	3.77%	6.00%	2,259.99	2,712.69
ARMA(2,2)$_{12}$	1,509.22	1,938.23	1.92%	2.04%	1,942.48	2,533.25
Exponential smoothing method	1,565.54	1,547.88	2.01%	1.77%	2,003.87	2,193.16

Table 4. Error Indicators for the Four Forecasting Methods.

D. Results

In this study, mean of absolute error (MAE), mean of absolute percent error (MAPE), and root-mean-square error (RMSE) were used as error indicators. The definitions for these error indicators are provided as follows:

$$MAE = n^{-1} \sum_{t=1}^{n} | F_t - Y_t |$$

$$MAPE = 100 n^{-1} \sum_{t=1}^{n} \left| \frac{F_t - Y_t}{Y_t} \right|$$

$$RMSE = \left(n^{-1} \sum_{t=1}^{n} (F_t - Y_t)^2 \right)^{0.5}$$

where n denotes the number of months to be forecasted (n = 12 in this case), Y_t represents the true product demand for month t of 2013, and F_t is the forecasted product demand for month t. The fixed-effect term in the linear model was compared with that in the linear mixed-effect model. As shown in Table 3, the absolute values of the coefficients for the explanatory variables in the linear mixed-effect model containing the random-effect term are greater (i.e., further from 0) than all of those in the linear model except for April. In addition, the standard errors and P values for all of the explanatory variables in the linear mixed-effect model are smaller than those in the linear model. Regarding the linear fixed-effect model, compared with January in a given year, the product demand was significantly greater in May and November (P value < 0.1), in July (P value < 0.05), and in March, April, and September (P value < 0.01). Compared with the linear fixed-effect model, the linear model yielded less significant results. The linear model is suitable for data containing mutually independent observation values. In this case, the observation values for product demand were correlated over time, thereby violating the assumption of the linear model. Therefore, the standard errors and P values for the linear model (Table 3) are not valid estimates, whereas those for the linear mixed-effect model are more reliable. Table 4 shows the error indicators for the four forecasting methods. Because this case involved three error indicators for each of the 20 product types, Table 4 presents the mean and standard deviation of the three error indicators. As shown in Table 4, the means and standard deviations of MAE, MAPE, and RMSE for the linear mixed-effect model are lower than those for the linear, ARMA, and exponential smoothing models, indicating that, in this case, the linear mixed-effect model is superior to the other three models. Regarding the model comparison (Table 5), the predictive values obtained through using the linear model to process the correlated data are unbiased [23]. However, the linear mixed-effect model (8) contains the random-effect term, whereas the linear model (9) does not. Therefore, in Model (8), the intercept and year-squared terms differ according to the product type, and thus the corresponding intercept values and coefficients differ based on the product type. In Model (9), the covariate of product type is not included in the explanatory variables, which generates identical predictive values for various product types in the same years and months. Thus, this model cannot predict the product demand for the individual product types, rendering its forecasting effectiveness inferior to that of Model (8). Regarding the exponential smoothing method, we considered product type and month as crucial influential factors, which were used as the basis for dividing the data into 240 data sets. For each product type, the pre-2012 monthly data were used to forecast the monthly product demand for 2013. In this manner, the exponential smoothing method was applied 12 times for each of the 20 product types. In addition, less than six observations from the historical data were used in the exponential smoothing method (for a given month, there were at most 6 sets of data from 2007 to 2012); consequently, the risk of inferential error was high because only a few observations were involved in the prediction. Regarding the seasonal time-series model ARMA(2,2)$_{12}$, we considered product type as a crucial influential factor and divided the data into 20 data sets according to product type. For each product type, 64 observations were used on average. The ARMA(2,2)$_{12}$ model was used to forecast the product demand for each product type by considering the correlation between the data for every 12 month. For both the exponential smoothing method and the ARMA(2,2)$_{12}$ model, the data were divided into subsets according to the product type and then used to estimate the monthly effect of each product type. Accordingly, although such procedure could consider the various monthly effects for various product types and the interaction between product type and month, it reduces the number of data observations involved in the prediction. In the linear mixed-effect model, 1295 data observations were used to estimate the random effect for each product type. The number of data observations used in the linear mixed-effect model was considerably more than that used in the exponential smoothing and time-series models, which could explain

	Number of models	Number of samples	Consideration for the effect of product type	Consideration for the effect of month	Consideration for the interaction effect of product type and month
Linear mixed-effect model	1	1,295	○	○	※1
Linear model	1	1,295		○	
ARIMA(2,2)	20	≤ 72	○	○	○
Exponential smoothing method	240	≤ 6	○	○	○

Table 5. Comparison of the Four Models.
※1 This effect is non-significant

why the linear mixed-effect model produced lower forecast errors. In addition, in Model (8), the random effect of the interaction term for month and year-squared term was considered and the likelihood ratio test was employed to examine whether this term is significant to this model. The results showed that only the random effects of the intercept and year-squared terms were significant, and the random effect of the month term did not significantly enhances its explanatory power for the data. Therefore, the random effect of the interaction term was not included in Model (8).

V. DISCUSSION

In summary, when applying the linear mixed-effect model, all of the historical data were used in one model to predict the monthly product demand for each product type, and to avoid problems resulting from dividing the data into smaller data sets. In this case study, using the linear mixed-effect model enables manufacturers who adopt the MTS production strategy to predict the amount of inventory they should stock. Furthermore, the model is more effective in forecasting product demand than is the time-series, exponential smoothing, and linear models.

Similar to the linear model, the linear mixed-effect model is typically used to examine the relationship between explanatory and research variables. Unlike the linear model, which assumes the observation values to be mutually independent, the linear mixed-effect model is suitable for examining correlated data. Because the data pertaining to business operations are generally correlated over time, the linear model is limited in applicability. By contrast, the linear mixed-effect model was initially developed to handle correlated data. Other methods such as the time-series and exponential smoothing methods formulate the correlation between observation values as parameters, and then estimate the parameters by data and forecast the observations by the estimates. When the time-series and exponential smoothing models were first developed, these methods were not aimed at analyzing the relationship between explanatory and dependent variables. Wang [19] proposed an exponential smoothing method that included explanatory variables and can be used to explore the association of research variable. Because this method is a relatively new development, most of statistical software packages have not yet incorporated related functions, and thus this method has not been widely used. By contrast, the linear mixed-effect model was developed more than 30 years ago, and related functions have been included in various statistical software packages. Using linear mixed-effect, time-series, and linear models to forecast product demand can yield negative predictive values. This phenomenon occurs when the linear mixed-effect model is used because $\boldsymbol{\varepsilon}_i$ in (2) is assumed to be normally distributed and the link function is an identity function. Negative values are usually obtained from historical data where product demand is zero or very low. To prevent this, predictive value was truncated at 0 (i.e., $F_t = \max(\hat{Y}_t, 0)$, where $\hat{Y}_t$ denotes a predictive value derived from any method, and F_t denotes an actual predictive value obtained from any prediction method). In other words, if $\hat{Y}_t > 0$, then $F_t = \hat{Y}_t$; if $\hat{Y}_t \le 0$, then $F_t = 0$. Some link functions in generalized linear mixed-effect model can deal with the case where dependent variable is restricted to $\hat{Y}_t \ge 0$ [44]. However, the prediction intervals for the random-effects in linear mixed-effect model are well developed [45, 46, 47, 48, 49]. It is useful to apply the prediction intervals in business operations for knowing whether the random-effect exists.

Implementing an MTS production strategy can enhance the competitive advantages of a manufacturer, enabling the manufacturer to rapidly satisfy product demand, thereby reducing internal and external transaction costs for handling orders. Employing this strategy also enables high batch centralized production and thus can reduce production costs and assist manufacturers in negotiating with material suppliers about the cost of materials. Because this approach enables short delivery times, customer satisfaction can be improved, thus attracting potential customers who need products immediately. Consequently, market share can be increased. MTS production also enhances the usage rate of production equipment. Companies that adopt an MTS strategy require an accurate forecasting method to realize these advantages. This study proposed an accurate forecasting method for determining the stock levels a company should determine for adopting the MTS production strategy, a topic that has seldom been discussed in studies on MTS production.

Using an MTS production strategy involves the potential risk of increasing inventory costs. Therefore, future studies should adequately apply the strengths of the linear mixed-effect model (e.g., accurately forecasting demand for multiple product types in one go) when forecasting. Future studies should consider investigating whether the forecasting intervals of the linear mixed-effect model can be coupled with various inventory strategies to assist manufacturers with adopting the MTS production strategy in order to develop an optimal business operation model in terms of optimal inventory time points and minimal inventory costs. In addition, to remain competitive, companies should enhance their organizational capability for elevating the threshold that enables competitors to develop similar operating models. Future studies are also recommended to explore the benefits that the MTS production strategy involving a linear mixed-effect model brings to the

various departments of an enterprise and the effects of such strategy on customer satisfaction and loyalty.

REFERENCES

[1] P. Danese and M. Kalchschmidt, "The role of the forecasting process in improving forecast accuracy and operational performance," International Journal of Production Economics, vol. 131, no. 1, pp. 204-214, 2011.

[2] I. J. Adan and J. Van der Wal, "Combining make to order and make to stock," Operations-Research-Spektrum, vol. 20, no. 2, pp. 73-81, 1998.

[3] T. Williams, "Special products and uncertainty in production/inventory systems," European Journal of Operational Research, vol. 15, no. 1, pp. 46-54, 1984.

[4] W. M. Guillot, "Strategic leadership: Defining the challenge," Air and Space Power Journal, vol. 4, 2003.

[5] R. R. Magee, "Strategic leadership primer," DTIC Document, Tech. Rep., 1998.

[6] S. A. Shambach, "Strategic leadership primer," DTIC Document, Tech. Rep., 2004.

[7] R. Blackburn, K. Lurz, B. Priese, R. Gob, and I.-L. Darkow, "A predictive analytics approach for demand forecasting in the process industry," International Transactions in Operational Research, 2014.

[8] C. A. Soman, D. P. Van Donk, and G. Gaalman, "Combined make-to-order and make-to-stock in a food production system," International Journal of Production Economics, vol. 90, no. 2, pp. 223–235, 2004.

[9] W. Popp, "Simple and combined inventory policies, production to stock or to order?" Management Science, vol. 11, no. 9, pp. 868–873, 1965.

[10] L. C. Hendry and B. Kingsman, "Production planning systems and their applicability to make-to-order companies," European Journal of Operational Research, vol. 40, no. 1, pp. 1–15, 1989.

[11] S. Rajagopalan, "Make to order or make to stock: model and application," Management Science, vol. 48, no. 2, pp. 241–256, 2002.

[12] M.-K. Chen and A. Cheng, "The study of supply chain inventory strategy under bto production environment," Journal of the Chinese Institute of Industrial Engineers, vol. 20, no. 4, pp. 398–410, 2003.

[13] L. Breiman., "Statistical modeling: The two cultures (with comments and a rejoinder by the author)," Statistical Science, vol. 16, no. 3, pp. 199–231, 2001.

[14] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," International Journal of Forecasting, vol. 20, no. 1, pp. 5–10, 2004.

[15] J. F. Muth, "Optimal properties of exponentially weighted forecasts," Journal of the american statistical association, vol. 55, no. 290, pp. 299–306, 1960.

[16] F. Chen, J. K. Ryan, and D. Simchi-Levi, "The impact of exponential smoothing forecasts on the bullwhip effect," Naval Research Logistics, vol. 47, no. 4, pp. 269–286, 2000.

[17] S. C. Graves, "A single-item inventory model for a nonstationary demand process," Manufacturing & Service Operations Management, vol. 1, no. 1, pp. 50–61, 1999.

[18] R. Gob, K. Lurz, and A. Pievatolo, "Electrical load forecasting by exponential smoothing with covariates," Applied Stochastic Models in Business and Industry, vol. 29, no. 6, pp. 629–645, 2013.

[19] S. Wang, "Exponential smoothing for forecasting and Bayesian validation of computer models," Ph.D. dissertation, Georgia Institute of Technology, 2006.

[20] G. E. Box, "Time series analysis: Forecasting and control,," Time Series and Digital Processing, 1976.

[21] G. E. Box and G. C. Tiao, "Intervention analysis with applications to economic and environmental problems," Journal of the American Statistical association, vol. 70, no. 349, pp. 70–79, 1975.

[22] A. Pankratz, "Forecasting with dynamic regression models," Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, New York: Wiley, 1991, vol. 1, 1991.

[23] L. C. Hamilton and D. Press, Regression with graphics: A second course in applied statistics. Duxbury Press Belmont, CA, 1992, vol. 1, no. 1.

[24] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," International journal of forecasting, vol. 22, no. 3, pp. 443–473, 2006.

[25] P. Hiesl and J. G. Benjamin, "Estimating processing times of harvesters in thinning operations in maine," Forest Products Journal, 2014.

[26] M. L. Bourbonnais, T. A. Nelson, M. R. Cattet, C. T. Darimont, G. B. Stenhouse, and D. M. Janz, "Environmental factors and habitat use influence body condition of individuals in a species at risk, the grizzly bear," Conservation Physiology, vol. 2, no. 1, p. cou043, 2014.

[27] Y.-C. Tsai, Y.-W. Chiu, J.-C. Tsai, H.-T. Kuo, S.-C. Lee, C.-C. Hung, M.-Y. Lin, S.-J. Hwang, M.-C. Kuo, and H.-C. Chen, "Association of angiopoietin-2 with renal outcome in chronic kidney disease," PloS one, vol. 9, no. 10, p. e108862, 2014.

[28] T. G. Holmes, W. M. Tonn, C. A. Paszkowski, and G. J. Scrimgeour, "Effects of winter surface aeration on pelagic zooplankton communities in a small boreal foothills lake of alberta, canada," Journal of Freshwater Ecology, no. ahead-of-print, pp. 1–14, 2014.

[29] N. Mohsin, G. Mourad, M. Faure, I. Szawarc, and J. Bringer, "Metabolic syndrome performs better than the individual factors in predicting renal graft outcome," in Transplantation proceedings, vol. 45, no. 10. Elsevier, 2013, pp. 3517–3519.

[30] A. Schwalm, Y.-S. Feng, J. Moock, and T. Kohlmann, "Differences in eq-5d-3l health state valuations among patients with musculoskeletal diseases, health care professionals and healthy volunteers," The European Journal of Health Economics, pp. 1–13, 2014.

[31] Y. Acar and E. S. Gardner, "Forecasting method selection in a global supply chain," International Journal of Forecasting, vol. 28, no. 4, pp. 842–848, 2012.

[32] P. G. Allen and B. J. Morzuch, "Twenty-five years of progress, problems, and conflicting evidence in econometric forecasting. what about the next 25 years?" International Journal of Forecasting, vol. 22, no. 3, pp. 475–492, 2006.

[33] R. Fildes, "The forecasting journals and their contribution to forecasting research: Citation analysis and expert opinion," International Journal of forecasting, vol. 22, no. 3, pp. 415–432, 2006.

[34] J. C. Pinheiro and D. M. Bates, Mixed-effects models in S and S-PLUS. Springer Science & Business Media, 2000.

[35] B. T. West, K. B. Welch, and A. T. Galecki, Linear mixed models: a practical guide using statistical software. CRC Press, 2014.

[36] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," Biometrics, pp. 963–974, 1982.

[37] N. Laird, N. Lange, and D. Stram, "Maximum likelihood computations with repeated measures: application of the EM algorithm," Journal of the American Statistical Association, vol. 82, no. 397, pp. 97–105, 1987.

[38] M. J. Lindstrom and D. M. Bates, "Newton Raphson and EM algorithms for linear mixed-effects models for repeated-measures data," Journal of the American Statistical Association, vol. 83, no. 404, pp. 1014–1022, 1988.

[39] P.-F. Pai, K.-P. Lin, and J.-S. Wang, "Stock price forecasting in Taiwan using ellipsoidal fuzzy system," Journal of the Chinese Institute of Industrial Engineers, vol. 21, no. 2, pp. 146–155, 2004.

[40] W.-Y. Hwang and J.-S. Lee, "A new forecasting scheme for evaluating long-term prediction performances in supply chain management," International Transactions in Operational Research, vol. 21, no. 6, pp. 1045–1060, 2014.

[41] Y. Cao, G. Wan, and F. Wang, "Predicting financial distress of Chinese listed companies using rough set theory and support vector machine," Asia-Pacific Journal of Operational Research, vol. 28, no. 01, pp. 95–109, 2011.

[42] F. B. e Silva, E. Koomen, V. Diogo, and C. Lavalle, "Estimating demand for industrial and commercial land use

given economic forecasts," PloS one, vol. 9, no. 3, p. e91991, 2014.

[43] X. Zhang, T. Zhang, A. A. Young, and X. Li, "Applications and comparisons of four time series models in epidemiological surveillance data," PloS one, vol. 9, no. 2, p. e88075, 2014.

[44] J. Jiang, Linear and generalized linear mixed models and their applications. Springer Science & Business Media, 2007.

[45] D. R. Cox, "Prediction intervals and empirical Bayes confidence intervals," Perspectives in Probability and Statistics, pp. 47–55, 1975.

[46] C. Morris, "Parametric empirical Bayes inference: theory and applications," Journal of the American Statistical Association, vol. 78, no. 381, pp. 47–55, 1983.

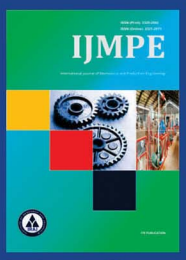
[47] R. Basu, J. K. Ghosh, and R. Mukerjee, "Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics," Statistics & probability letters, vol. 63, no. 2, pp. 197–203, 2003.

[48] S. Chatterjee, P. Lahiri, and H. Li, "On small area prediction interval problems," The Annals of Statistics, vol. 36, pp. 1221–1245, 2008.

[49] M. Yoshimori, and P. Lahiri, "A second-order efficient empirical Bayes confidence interval," The Annals of Statistics, vol. 42, no. 4, pp. 1–29, 2014.

★★★

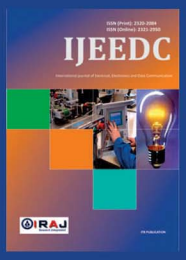
IRAJ INTERNATIONAL JOURNALS



IJMPE

www.ijmpe.iraj.in

International Journal of Mechanical and Product Engineering
ISSN(P):2320-2092
ISSN(e):2321-2071



IJEEDC

www.ijeedc.iraj.in

International Journal of Electrical, Electronics and Data Communication
ISSN(P):2320-2084
ISSN(e):2321–2950

IJACEN

www.ijacen.iraj.in

International Journal of Advance Computational Engineering and Networking
ISSN(e):2320-2106
ISSN(P):2320-2106

Indexing Partners

International Journal of Soft Computing And Artificial Intelligence (IJSCAI)
ISSN(P):2321–404X
ISSN(e):2321–4384
www.ijscai.iraj.in

International Journal of Advances in Computer Science and Cloud Computing (IJACSCC)
ISSN(P):2321–4058
ISSN(e):2321–4392
www.ijacscc.iraj.in

International Journal of Advances in Science, Engineering and Technology(IJASEAT)
ISSN(P):2321-8991
ISSN(e):2321–9009
www.ijaseat.iraj.in

International Journal of Industrial Electronics and Electrical Engineering(IJIEEE).
ISSN(P):2347-6982
ISSN(e):2349-204X
www.ijieee.iraj.in

International Journal of Advances in Mechanical and Civil Engineering(IJAMCE)
ISSN(P):2394-2827
www.ijamce.iraj.in

International Journal of Management and Applied Science (IJMAS)
ISSN(P):2394-7926
www.ijmas.iraj.in

International Journal of Advance in Electronics and Computer Science (IJAECS)
ISSN(P):2393-2835
www.ijaecs.iraj.in

IRAJ Journals Listed in University Library
MIT, University of California Berkeley, Stanford, Cambridge, Oxford, Harvard
Visit for Upcoming Conferences - www.ieeeconference.com

Digital Xplore

This book will be available online at DigitalXplore
www.digitalxplore.org

ISBN
978-93-90150-34-2