



Foundation University
Islamabad



FUSST INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND ENGINEERING 2024

Proceedings

FICSE



FUSST International Conference On Computer Science and Engineering

Email: ficse@fui.edu.pk

Web: <http://ficse.fui.edu.pk>

Foundation University School of Science & Technology



**Proceedings of
FUSST International Conference on Computer
Science and Engineering**

FICSE 2024

**Foundation School of
Science and Technology,
New Lalazar, Rawalpindi
Pakistan**
Email: ficse@fui.edu.pk

Executive Committee

Maj Gen Muhammad Kaleem Asif (Retd)
Foundation University Islamabad
Patron in Chief

Prof. Dr. Abdul Ghafoor
Foundation University Islamabad
Patron

Prof. Dr. Muhammad Shaheen
Foundation University Islamabad
Conference Chair

Dr. Shariq Hussain
Foundation University Islamabad
Conference Co-Chair

Technical Program Committee

Prof. Dr. Muhammad Shaheen
Foundation University Islamabad

Dr. Shariq Hussain
Foundation University Islamabad, Pakistan

Dr. Sajid Ali Khan
Foundation University Islamabad

Dr. Engr. Abdur Rehman
Foundation University Islamabad

Dr. Ateeq Ur Rehman
Foundation University Islamabad

Dr. Tehmina Karamat Khan
Foundation University Islamabad

Dr. Arif Jamal Malik
Foundation University Islamabad

Dr. Muhammad Haneef
Foundation University Islamabad

Dr. Imran Daud
Foundation University Islamabad

Dr. M. Aqeel Iqbal
Foundation University Islamabad

Dr. Maryam Kausar
Foundation University Islamabad

Table of Contents



FUSST International
Conference on Computer
Science and Engineering

FICSE

Research Article

A Business Intelligence System for Used Vehicle Trading (BIS-VeT)

Engr. Dr. Umar Mahmud, P.E.

Futureaware Tech (PVT) LTD Islamabad, Pakistan
(umar.mahmud@futureaware.tech).

Article Citation:

Mahmud et al. (2024). "A Business Intelligence System for Used Vehicle Trading". *FUSST International Conference on Computer Science and Engineering*

Abstract:

Business Intelligence (BI) systems are used to identify opportunities and plot trends based on data gathered from multiple sources. A BI system uses Extract, Transform, and Load (ETL) process to acquire data of interest from multiple sources. This data can be used to draw insights as well as calculate long-term probabilities. BIS-VeT connects with two different online vehicle trading services and acquires posts as data. This data is then cleaned to get price quotes for different vehicle types. Visualizations are drawn to see the price trend of multiple vehicles. Furthermore, the long-term probability of trade popularity of a vehicle type is calculated by compiling a Markov Chain. This system is useful for customers, businesspersons, and car enthusiasts alike. The long-term probability calculated based on the data show the popularity of Corolla GLI higher than that of Honda City as a resale vehicle in Pakistan.

Keywords: Business Intelligence, Data Science, Markov Chain, Used Vehicle Trading.

1. Introduction

The emergence of Business Intelligence (BI) has seen many advancements in healthcare, social sciences, trading, and other data science domains [1]. BI systems are useful for decision support systems that have large data from heterogeneous sources [2]. Decision support requires the effective use of business analytics, statistical techniques, and plotting trends of numerical data. BI system uses statistical and probabilistic techniques to learn future states and directions which is useful for e-commerce applications [3]. The prediction has been done using Bayesian Probability and Markov Chains and has been utilized in a wide variety of domains [3]. At the heart of a BI system is an analytics module that can use Machine Learning based algorithms to classify unknown states [4].

Data science is an ensemble concept that includes statistics, probability, data engineering, machine learning, and visualization [7]. Data science now enables companies to utilize information to make better decisions and identify opportunities. The power of this domain can be utilized by small domains like smart homes, and communities, as well as large city-wide domains for service provisioning and the general welfare of the masses [8], [9]. Data science when employing machine learning techniques in business processes becomes BI. The major contribution of a BI system is identifying challenges in production, manufacturing, transactional data, and mundane tasks. A BI system requires the representation of knowledge and the extraction of information of interest. The data can be large and may require a data warehouse for storage and subsequent processing. Using machine learning tools can help identify opportunities as well as answer specific questions of interest. In today's world, with the increase in population transportation using fossil-fueled vehicles is becoming a constraint. Vehicle manufacturers cannot compete with the growing demands of vehicle buyers, and there is a large market for buying and selling used vehicles. Globally the used car market size is 1.5 trillion USD in 2022 [5]. The number of car trades in developing nations is much higher than in developed nations with around 14 million used vehicles sold in developing countries in Africa between 2015-2018 [6]. Car trading is carried



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

out using online portals where buyers and sellers can connect. These portals are free to use and provide services to both buyers and sellers. The data records are large and there is no interoperability among such web applications. Potential users must look for their requirements and post their queries on multiple platforms. Furthermore, a user cannot work out the trends and future probabilities from these systems. The novelty of this system is based on the use of VI system coupled with Markov Chain to ascertain the long-term probability of vehicle resale market. Furthermore,

this system is designed to incorporate new data every month. The time period can be reduced to daily or weekly basis as well as increased to quarterly data which is useful for brand managers and sales executives.

Based on contemporary online car trading portals, and their lack of interoperability and BI support, the following research questions are identified.

- RQ I: How can contemporary systems be integrated for BI support for car traders?
- RQ II: How can long-term probabilities of trade popularity be calculated for different vehicle types?

The rest of the paper is organized as follows. Section II outlines the related work, while Section III presents the design of the BIS-VeT. Section IV discusses the results, and the paper concludes in Section V.

1. RELATED WORK

Eskandar highlights that BI and trading technologies go hand in hand [7]. Manufacturers, traders, and consumers can utilize the power of trading technology to increase their revenue.

Since market trading occurs daily, it can be viewed as a random variable dependent on time. This makes market trading a stochastic process [8].

Mettle et al., have viewed share prices as a stochastic process and used Markov Chain to model them [9]. They show that the expected mean return for stock prices can be calculated based on historical information and have used weekly data from Ghana Stock Exchange to show the effectiveness of their technique. Langer and Lemoine have also used probability to estimate a market's probability in the event of uncertain events [10].

Chandra and He have used Bayesian neural networks to forecast stock prices and long-term probabilities. The authors have shown that stock prices are highly volatile and using Bayesian neural networks can provide reasonable predictions under uncertainty [11]. Zhao has developed an ensemble technique comprising Bayesian probability and regression models for stock market forecasting [12]. Pykes have used the ARIMA model to forecast sales data of Walmart [13].

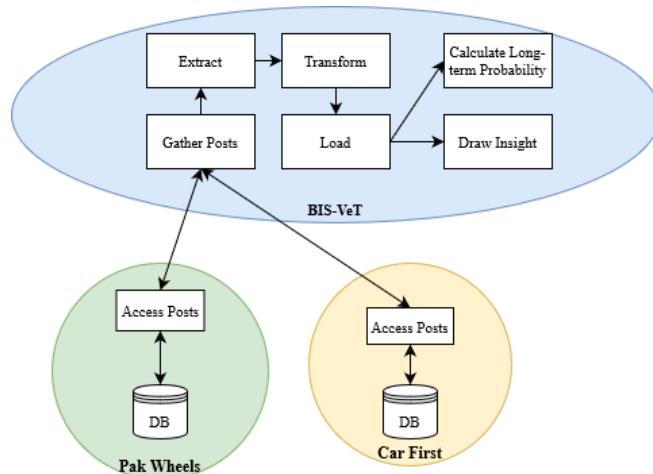


FIG. I

CONCEPT DIAGRAM OF BIS-VET

Multiple car trading portals are available for use by consumers. The portals and their features are compared with BIS-VeT, in Table I. It is pertinent to note that a BI system is designed for a Business-2-Business (B2B) model, while the applications that are available for public use are primarily Business-2-Consumer (B2C) models.

2. DESIGN OF BIS-VET

BIS-VeT is developed using a Service-Oriented Architecture (SOA) suitable for interacting with other car trading services. BIS-VeT uses HTTP to communicate with the services. For the proof of concept, the services selected for interaction are Car First and Pak Wheels. Both services are available online via the web, and data can be gathered using web scraping.

A BI system requires the implementation of Extraction, Loading, and Transformation (ELT) Processes. The gathered data is extracted and then loaded into an appropriate data structure. To normalize the data a transformation process is run. Now the data can be used to draw insights, plot trend lines, or simply be used to calculate long-term probabilities.

Fig. I shows the concept diagram of BIS-VeT. BIS-VeT communicates with other services to gather data that is available as posts. This data is then passed through the ELT process. This can then be used to draw insights as well as calculate long-term probabilities. The mechanism essentially creates a data warehouse through web scrapping and extraction in to the system. This is then subject to compilation of Markov Chain and the probability transition matrix which is then used to determine the long-term probability. Since the data is acquired at regular intervals, the mechanism is discrete and the resulting Markov Chain is a Discrete-Time Markov Chain (DTMC)

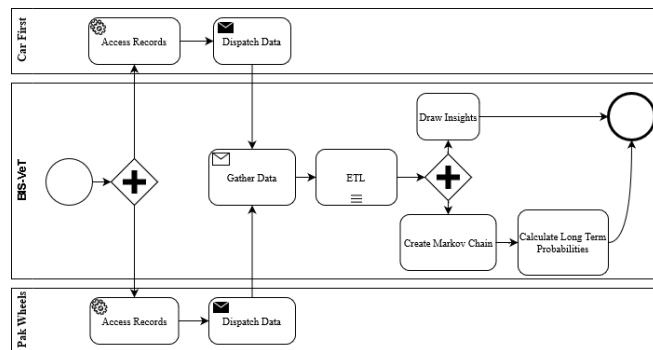


TABLE I. COMPARISON OF EXISTING SYSTEMS

System	URL	Feature		
		BI	Business Model	Forecasting
Car First	https://www.carfirst.com	✗	B2C	✗
PakWheels	https://www.pakwheels.com/	✗	B2C	✗
gari.pk	https://www.gari.pk/	✗	B2C	✗
BIS-VeT	-	✓	B2B	✓

Since BIS-VeT is envisaged as an SOA, it is suitable to use Business Process Modelling Notation (BPMN) to model the processes and their interactions across multiple business processes. Fig. II shows the BPMN diagram, where the interacting business includes Car First, Pak Wheels, and BIS-VeT. The businesses and their constituent processes are shown in the swim lanes of the BPMN diagram. The processes communicate sequentially via message passing. BIS-VeT gathers data using the HTTP GET method. Table II shows the processes in the BPMN diagram and their description. Both services are accessed concurrently to gather the latest data.

TABLE II. PROCESSES AND DESCRIPTION

Process	Swim Lane	Description
Access Records	Car First/Pak Wheels	This process uses web scrapping to access the posts.
Dispatch Data		This process gathers the posts and sends them to BIS-VeT. The data is sent in HTML and needs to be cleaned before use.
Gather Data	BIS-VeT	This process gathers data from online services and stores it in files. The data is then cleaned to remove unnecessary information and only posts are stored. The posts are stored in chronological order
ETL		This runs the Extract, transform, and Load process of a BI system executed sequentially. Extraction process requires that the data of interest be taken from the stored data e.g., the posts concerning Toyota Corolla GLi, since last 3 months. Transformation process requires that the data be transformed into a format suitable for representation. Since the data is acquired through multiple autonomous sources in the public

Process	Swim Lane	Description
		domain, the data must be transformed into a suitable format for use. Load process loads the data in appropriate data structures for processing.
Draw Insights		Drawing insights requires plotting the data and then identifying opportunities. The point of interest is plotting of price of a used car type.
Create Markov Chain		A Markov chain and an associated probability distribution matrix are compiled. This matrix can be updated after new data is acquired.
Calculate Long Term Probabilities		The long-term probability can be calculated using the probability distribution matrix of the Markov Chain. This allows us to make decisions based on the current state.

3. RESULTS

To verify the effectiveness of BIS-VeT, data from both online services are gathered. The point of interest is the price of used cars. To evaluate, the trend of two different used cars that are popular in Pakistan is selected over a period of two months. Fig. III plots the car price of each car type acquired from posts on both services. Since the posts can be placed anytime, the weekly average car price is plotted in Fig. III. Table III shows the data of the point of interest. The cars selected are Toyota Corolla Gli, Model 2005-2009, and Honda City1.3, Model 2005-2009.

TABLE III. AVERAGE PRICE PER WEEK

Week	Price (Millions of PKR)	
	Toyota Corolla Gli Model (2005-2009)	Honda City 1.3, Model (2005-2009)
1st Week of June	1.15	1.45
2nd Week of June	1.21	1.52
3rd Week of June	1.22	1.53
4th Week of June	1.22	1.48
1st Week of July	1.21	1.44
2nd Week of July	1.32	1.46
3rd Week of July	1.34	1.49
4th Week of July	1.31	1.47

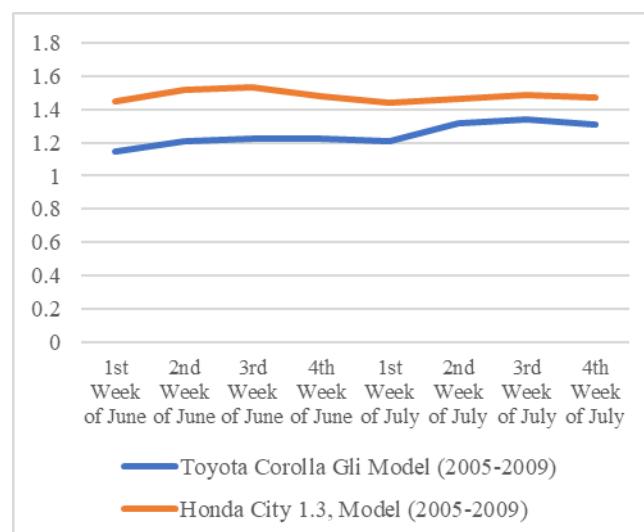


FIG. III PLOT OF AVERAGE PRICE PER WEEK

The plot shows the price trend of both car types. The trend can be used through regression to plot future values. Given that a buyer wants to invest in a car type and has a specific budget. This plot and its trend can be used to ascertain which car type a person with a budget can buy.

For car dealers, manufacturers, and enthusiasts the long-term probability of a popular car is of interest. Using the Markov chain, the long-term probability can be calculated easily. To measure the long-term probability, the count of legitimate posts of a car type must be maintained. In the period of interest as shown in Table III and Fig. III, there have been a total of 8,934 posts concerning the cars of interest. Out of these 6871 posts have been of Toyota Corolla Gli Model (2005-2009). The probability transition matrix based on this data is shown in Equation I.

$$P = \begin{bmatrix} 0.77 & 0.23 \\ 0.23 & 0.77 \end{bmatrix} \quad (I)$$

The long-term probabilities for both types of cars are given in Equations II and III.

$$p(\text{popularityOfToyotaCorollaGli}) = 0.57 \quad (II)$$

$$p(\text{popularityOfHondaCity1.3}) = 0.43 \quad (III)$$

Equations II and II confirm that Toyota Corolla Gli is a popular brand and there is a higher probability of its trade through online services when compared with Honda City 1.3.

4. CONCLUSION

This paper presents BIS-VeT, a system that enables BI capabilities for online car trading services. Trade of used cars is a stable business in third-world countries. Many online services are available that are used by customers to buy and sell cars. Since these online services are specific, a customer must gather information through multiple applications. Furthermore, there is a need to plot trends and long-term probabilities of the popularity of a vehicle.

BIS-VeT is designed as an SOA, that gathers data from posts on two different online services. The data is then subject to the ETL process and can be used to draw insights as well as create visualizations. The concept and a BPMN model highlighting the processes are presented.

The data is then used to plot price trends, based on the posts on multiple services. The price trend is useful to monitor the business and can be extrapolated for future values. However, a person on a budget needs to know the popularity of a car type as well. This popularity helps in understanding the traceability of a used car. The long-term popularity probability can be calculated using a Markov chain based on the number of posts of a car type.

The research question posed in Section I are answered here:

- RQ I: How can contemporary systems be integrated for BI support for car traders?

Contemporary systems can be integrated, and data can be acquired as well as processed using ELT. This data can then be used to draw insights, visualization, and find opportunities.

- RQ II: How can long-term probabilities of trade popularity be calculated for different vehicle types?

Long-term probabilities of trade popularities can be calculated using the Markov chain. A Markov chain is compiled by counting the number of posts. This can then be used to calculate long-term probabilities.

Future work requires extensive testing and development of the software product. Furthermore, the long-term behavior probability of price range and other features of interest can also be measured using the Markov chain.

5. ACKNOWLEDGMENT

A software product based on BIS-VeT is under development in FUI by Ms. Laila Bukhari, Ms. Sana Fatima, and Mr. Abdul Raheem. The product is envisaged to be completed in mid-2023.

6. REFERENCES

- [1] F. Provost and T. Fawcett, Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly Media, Inc, 2013.
- [2] B. Marr, Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things, Kogan Page, 2017.
- [3] X. Ye and M. Jonillo, "The application of business intelligence system based on big data in e-commerce data analysis," *Journal of Computing and Electronic Information Management*, vol. 12, no. 1, 2024.
- [4] S. Hussain, U. Mahmud and S. Yang, "Car e-Talk: An IoT-enabled Cloud-Assisted Smart Fleet Maintenance System," *IEEE Internet of Things*

Journal, pp. 1-1, 2020.

- [5] T. Mitchell, Machine Learning, McGraw-Hill, 1997.
- [6] L. Cao, "Data Science: A Comprehensive Overview," *ACM Computing Surveys*, vol. 50, no. 3, pp. 1-42, 2018.
- [7] U. Mahmud, S. Hussain, A. Sarwar and I. K. Toure, "A Distributed Emergency Vehicle Transit System Using Artificial Intelligence of Things (DEVeTS-AIoT)," *Wireless Communication and Mobile Computing*, vol. 2022, pp. 1-12, 2022.
- [8] U. Mahmud, S. Hussain, A. J. Malik, S. Farooqui and N. A. Malik, "Realizing IoE for Smart Service Delivery: Case of Museum Tour Guide," in *Smart Systems Design, Applications, and Challenges*, IGI Global, 2020.
- [9] GrandViewResearch, "Used Car Market Size & Share Report, 2022-2030," 2021. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/used-car-market>. [Accessed 13 12 2022].
- [10] J. Arvin, "Cars too dangerous and dirty for rich countries are being sold to poor ones," 29 10 2020. [Online]. Available: <https://www.vox.com/21536763/climate-change-africa-un-used-cars-environment>. [Accessed 13 12 2022].
- [11] A. Eskandar, "Business Intelligence and Trading Technology, I Wed Thee," 25 11 2013. [Online]. Available: <https://www.marketsmedia.com/business-intelligence-trading-technology-wed-thee/>. [Accessed 13 12 2022].
- [12] S. M. Ross, Introduction to Probability Models, 11 ed., Academic Press, 2014.
- [13] F. O. Mettle, E. N. B. Quaye and R. A. Laryea, "A methodology for stochastic analysis of share prices as Markov chains with finite states," *Springerplus*, vol. 3, no. 657, pp. 1-11, 2014.
- [14] A. Langer and D. Lemoine, "What Were the Odds? Estimating the Market's Probability of Uncertain Events," *NBER Working Paper Series*, pp. 1-43, 2020.
- [15] R. Chandra and Y. He, "Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic," *PLOS ONE*, vol. 16, no. 7, pp. 1-32, 2021.
- [16] X. Zhao, "Application of Bayesian Regression Model in Financial Stock Market Forecasti," in e 2nd International Conference on Management, Economy a, Kaula Lumpur, 2021.
- [17] K. Pykes, "Forecasting with Stochastic Models," 12 6 2020. [Online]. Available: <https://towardsdatascience.com/forecasting-with-stochastic-models-abf2e85c9679>. [Accessed 13 12 2022].



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:

Khan et al. (2024). "Selecting discriminative features for scene structure learning by combining multi-layered convolutional features and handcrafted features". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

level features extraction techniques. These methods need pre-processing steps and a pre-defined map for features extraction. Other approaches [5, 11, 13, 17-24] utilize the Bag of Visual Words (BoVW) technique [7] for scene understanding and recognition task. Classification results of these approaches go down when the scene structure becomes more complex, which increases the chance of intra-class variations [14].

Recently, convolutional neural networks (CNN) based approaches [25-28] illustrate high performance on ImageNet dataset [29] once they use up to 80% images for training and 20% images for testing purposes. Therefore, CNN become the backbone of many types of research, including object recognition [30, 31], object detection, and action

Selecting discriminative features for scene structure learning by combining multi-layered convolutional features and handcrafted features

Altaf Khan^a, Eid Rehman^b,

^aDepartment of Software Engineering, University of Mianwali, Pakistan.

^bDepartment of Computer Science & IT, University of Mianwali, Pakistan.

Corresponding author: altaf.khan@umw.edu.pk, Eid.rehman@umw.edu.pk

Abstract:

Scene type identification from a single image is an overriding task in the area of the computer vision. It has a major role in different applications such as autonomous car driving, object localization, and 3D TV. However, it is a difficult task to identify the scene type when the foreground and background objects have complex relationships in a single scene. State-of-the-art methods address this issue by using convolutional neural networks (CNN) architecture. Some methods use both CNN and traditional handcrafted features to identify the scene type from a single image. These methods show weak performance for scene identification because of features redundancy and lack of discriminative features set of the medium size input database. Selecting discriminative features from the combined CNN and handcrafted features and use of these features for scene representation is proposed in this manuscript. It minimizes redundancy in the features set while also improving the accuracy of scene type identification. To obtain the discriminative features of the image scene, local and global features of CNN at multi-layered (intermediate and final layers) are extracted and combined with handcrafted features (HF) into a single vector. The HF are extracted from the same image by dividing it into $n \times n$ parts. Then the features selection (FS) technique is applied to get the discriminative feature vector for learning the scene structure from a single image. The proposed method, CNN-HF-FS, is applied to the 12000 images of the medium-scale dataset and achieves 93.62% recognition accuracy, which is higher than 11.37%, 11.74%, and 1.08% from the GoogleNet, ResNet-50, and HF-MSF-HFF [16] methods, respectively.

Keywords: CNN, Feature Selection Techniques, Multi-layered CNN, Handcrafted Features, Scene Representation, Scene Geometry

1. Introduction

Real 3D world around us can be captured in 2D images [1], which is called an image scene. Human eyes can view ten Gigabytes of data per second [2] and understand the impression of a scene depth from an image. Developing such a system for a machine that identifies the same impression as a human from a scene image is a challenging task for researchers [3]. Mainly, the scene can be categorized into two parts: indoor and outdoor scenes [4-8]. The outdoor scene consists of ground-sky, ground-background and sky, ground-background, mountain-sky, etc. Most of the indoor scenes are related to human-made objects, e.g., rooms, kitchens, offices, corridors, etc. The outdoor scenes are comparatively less complex than the indoor scenes because of the smaller number of background objects. Several researchers have divided these scenes into limited categories to achieve a well-rounded classification based on the image scene structure. Categorize these scenes into 12 different classes, also called scene geometry. Scene recognition is very useful for multiple applications of computer vision, e.g., 3D Television, autonomous car, video understanding and localization [4-8][4, 9][4, 9, 10]. Although various researchers [9-13] have proposed scene understanding and object-to-scene relationship classification techniques but due to the complex scene structure's intra-class diversity and inter-class homogeneity, it is still a difficult process [14-16]. The traditional approaches, such as [4] and [5, 9, 10] utilize handcrafted low-

recognition [32-34]. CNN pyramid mainly consists of Convolutional, pooling, and fully connected (FC) layers. It was trained using backpropagation methods on large datasets, and it exhibited significant accuracy in classification tasks including scene categorization, face identification, and object localization [26, 28, 35-38]. However, classical CNN models show low performance on medium size datasets because they may lose much local information at the global-level (end layers) during many convolutional and pooling operations [14]. Thus, in some research works [14-16, 39-41], the deep features at multiple convolutional layers are extracted and utilized for scene recognition, and significant results have been achieved for scene recognition. Most of these approaches utilize the features fusion strategy in which multi-layer features are combined at the final stage and then the classifier is applied for the identification of scene type, e.g., [14]. These methods do not show high performance for scene recognition because of feature redundancy and lack of discriminative features for the medium-sized scene dataset. Selecting discriminative features from the combined CNN and handcrafted features and these features are used for scene understanding is proposed in this article. It minimizes redundancy in the feature set while also improving the accuracy of scene-type prediction. To capture the discriminative features of the input scene, the local-level features at intermediate layers and global-level features at the final layers of CNN architecture are extracted and combined with handcrafted features (HF) into a feature vector. The HF is extracted from the $n \times n$ parts of a single image. Then the features selection (FS) technique is applied to get the discriminative feature vector for learning the scene structure from a single image. The proposed CNN-HF-FS method is applied to the medium-size 12-scene dataset [6] and it achieves 93.62% classification accuracy, which is robust compared to the existing methods [10, 14, 16, 42].

1.1. Motivation of study

The motivation of the paper is:

- Firstly, extract the multi-layer local and global features of the deep CNN model at different stages to keep the discriminative information of the scene structure and combine them with handcrafted traditional features at the last stage to improve the discriminative property of the scene for accurate prediction of the scene type.
- Secondly, the utilization of the feature selection algorithms for selecting the discriminative features from the combined set of features, reduces feature redundancy and complexity at the classifier level while showing improvement in the classification performance.

1.2. Organization of the study

The remaining manuscript is arranged as follows: Section 2 describes the related work. The proposed method is described in Section 3. Section 4 illustrates the implementation of the proposed method. Section 5 represents experimental results and discussion. Section 6 concludes the manuscript.

2. Related work

The related work is typically divided into two parts: Low and high-level features-based scene-type identification approaches. Low-level features are traditional features and are mostly represented in the form of texture, color, texture, or structure information to represent the image scene data locally. These properties are handcrafted by applying the man-made techniques, e.g., $n \times n$ grid parts of an input image. Oliva et al. [5], Hoiem et al. [43], Nedovic et al. [4], Lou et al. [9], and Khan et al. [10] extract Weibull distribution's parameters [44], colors, Perspective-line [45], and HOG [46] features as low-level from $n \times n$ parts of the input image to understand the scene structure. These algorithms use Support Vector Machine (SVM) to update the parameters for scene structure learning and achieve 38.0%, 47.30 %, and 82.50% performance on their mentioned scene datasets. The implementation of these methods is hard and based on the traditional approaches of feature extraction that need a large number of pre-processing steps including image segmentation and features extraction techniques.

Mid-level features, BoVW [7] represents the semantic bridge between the high and low-level features. Some researchers carried on to develop by combining the semantic information, e.g., J. Sánchez et al. [11] used Scale Invariant Feature Transform (SIFT) in the framework of the Fisher Kernel (FK) with Gaussian mixture model (GMM) in the replacement of $n \times n$ parts and it illustrates the performance of 47.2% when the SUN397 dataset [47] is used with 50% training samples. However, these features show the low performance of classification tasks when complex scene images are given to the model [14]. Instead of this, high-level deep features of CNN architecture can obtain more fine and pixel-level information of the image. In 1989, Yann et al. [48] presented a multi-layer neuron-based network using BP calculation to learn the hand-composed structure. It initialized the new track trend in the CNN architecture. In contrast, it did not become popular quickly because of the unavailability of much equipment and large-scale label datasets for training the model. In the last decade, the recent CNN architectures, e.g., [25-27, 49, 50] have significantly improved classification tasks, become a well-liked framework in the field of artificial intelligence, and demonstrated cutting-edge performance on large image scene datasets. In recent scene classification research, some researchers use the deep convolutional features at multiple layers and the last (FC) layers for scene identification. The robust and effective typical CNN methods take only semantic data by activating the last layer that performs well for scene recognition task. However, when using a small or medium-scale dataset, it does not provide a

detailed description of the object and scene-to-object relationship during the multiple calculations on the convolutional and pooling layers [14, 15]. Tang et al. [14] introduced the G-MS2F method to study the middle layers characteristics of the GoogLeNet [28] model for scene representation and to address the intra-class variations problem of indoor scene images. On a dataset of 15 scenes with a standard setting, it obtains 92.90% recognition accuracy. As ResNet shows better performance compared to GoogLeNet model, Liu et al. [15] proposed a state-of-the-art ResNet-based model in which multi-layer features are combined for scene representation using softmax function. The author title this technique as ‘fine-tuning on transfer learning model (FTOTLM)’ in which Global average pooling (GAP) is applied to minimize the features dimensionality. On a 15-scene dataset, this method achieves the recognition accuracy of 94.04% and on the MIT67 and SUN397 datasets, it achieves accuracy of 74.63% and 65.46%, respectively. Khan et al. [16] introduce a novel approach for scene classification by integrating the rich handcrafted features with multi-layer deep CNN features (HF-MSF) using features and score-level fusion techniques. This method improves recognition accuracy of 12-scene dataset [6] to 92.54% and 93.96% when the ResNet and GoogLeNet architectures are used, respectively, by integrating all features into a single vector. It achieves 97.86% state-of-the-art results of scene representation on 15-scene dataset when the HF-MSF method uses feature-level at different stages and score-level fusion technique at the classifier-level. Similarly, it achieves 95.17% when the same setting is applied on the 12-scene dataset. However, integrating the features at multiple stages and applying score-level fusion at the classifier-level increases the computational cost of the HF-MSF method. It also increases the implementation complexity of the method as well. Some researchers, such as [42] obtain the deep features from the middle layers of both the inception and residual architecture and combine them in a CNN framework with a set of classifiers. These multi-layer intermediate stages or blocks contain the local structural detail while the last layer contains the semantic global structure of the image scene. In addition, the set of features at each block is used to predict the class score using a classifier. Finally, the class scores of the classifiers are combined to predict the class label. Besides this, related approaches including [30-34, 38, 51, 52] achieve good performance but they are designed for particular recognition issues such as human action recognition [32-34, 38], biomedical image task recognition [51, 52] and object identification [30, 31].

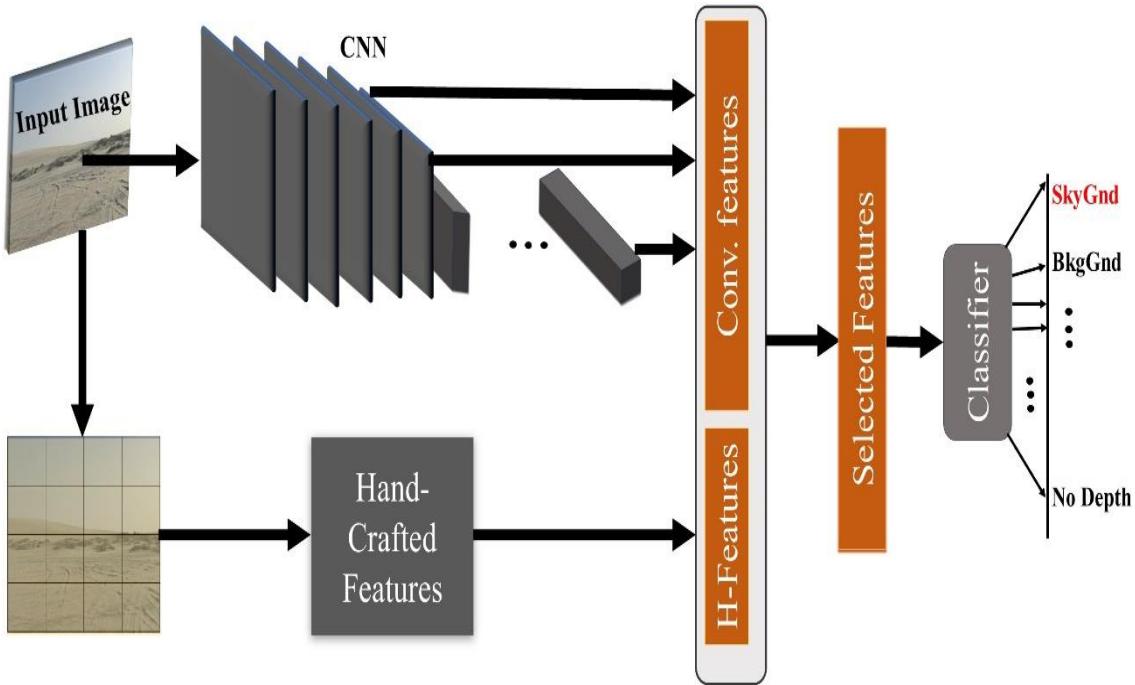


Fig 1. Proposed CNN-HF-FS Scene recognition method.

3. Proposed method

The new method, CNN-HF-FS is introduced to improve the discriminative features of image scene geometry structure. It uses the set of deep features from a standard CNN architecture combined with traditional features from $m \times m$ parts of the image as visualized in Fig 1 and described in Algorithm 1, and 2. Algorithm 1 and Fig 1 present the overall process of the proposed method. The figure has three main parts, first is the CNN model used for extraction of the convolutional features. The second part is the extraction of handcrafted features from the same image. In the third part, discriminative features are selected using feature selection algorithm. Finally, the Figure shows that a classifier is applied to predict the class type of the input image. Algorithm 2 particularly presents the extraction of the traditional features for each input image. In Algorithm 1, the multi-layered features, $F_j^k, j = 1, 2, 3 \dots N$, N is training sample, extracted from the k th stage, where $k=1, 2, 3, \dots, n$. F is a deep feature vector of the k th stage (see lines 1-4). The `Features_extracted_Using_DeepCNN(.)` is a function representing the deep CNN model including the activation

Algorithm 1: Proposed CNN-HF-FS Scene Recognition

Input: N training sample, T total number of images, Testing images $M=N+1$: T , n number selected stages of CNN, Y_N training and Y_M testing labels.

Output: Acc, Pr, Re, F-Sc

// Extracting deep and traditional features

- 1: **for** $j=1$ to T **do** // for each image I_j
- 2: **for** $k=1$ to n **do** // selected number of stages n of CNN
- 3: $F_j^k = \text{Features_extracted_Using_DeepCNN}(k, I_j)$ // The value of k is substituted in the CNN architecture.
- 4: **end for** // n stages
- 5: $F_j^k = L2_Norm(F_j^k)$ //using equation (1).
- 6: $F_j^{Trad} = \text{Traditional_Features_extraction}(I_j)$ // According to Algorithm 2, a set of traditional features of input image I_j .
- 7: **end for**
- 8: $F'_j = F_j^{Trad} + \cup_{k=1}^n \{F_j^k\}$ //concatenation of traditional and deep features.
- 9: $F''_j = FS(F'_j)$ // Feature selection algorithm.
- // Training part
- 10: $\bar{X} = \{(F''_j, Y_j)\}_{j=1}^N$ //labelling the features vector
- 11: $Train_Cl = \text{Training_Classifier}(\bar{X})$, //Training the classifier
- // Testing Part
- 12: **for** $j=1$ to M **do**
- 13: $L_j = \text{Classify}(Train_Cl, F''_j)$ predict label L_j of image j .
- 14: **end for**
- //Performance
- 15: [Acc, Pr, Re, F-Sc]=Calculate_Performance (L , Y_M). // L is a label vector of M images.

end algorithm

function, **Relu** and global average GAP, to extract the deep features at each stage. Then at line 5, these extracted features, F_j^k , from the k th stage are normalized by L2 norm equation, given by,

$$F' = \frac{F}{\sqrt{\|F\|_2^2 + \epsilon^2}}, \quad (1)$$

where ϵ is constant, F is input features, and F' is a normalized vector.

After that, at line 6 of algorithm 1, the low-level traditional features set are extracted using **Traditional_Features_extraction(I_j)** function which is described in algorithm 2. I_j is the input j th image and this function return F_j^{Trad} features vector for the j image. Algorithm 2, line 1 shows the size of each patch. Lines 1-15 represent feature extraction steps of texture gradient (lines 5-6), HOG (line 7), and color features (lines 8-13), which contain rich information of the image scene [4, 9, 10]. The line 16 represents the normalization of these features using equation (1). Then, line 17 shows the concatenation of these features into a single vector, F^{Trad} . In algorithm 1, F^{Trad} and deep features are concatenated into F'_j vector as shown in equation (2),

$$F'_j = F_j^{Trad} + \cup_{k=1}^n \{F_j^k\}, \quad (2)$$

where, $j = 1, 2, \dots, N$ are size of samples, '+' and \cup indicate as the concatenation operator. F'_j is a combined feature that is given as input to the FS technique. The k is the stage, where $k=1, 2, 3, \dots, n$.

At line 9, the algorithm is used to select the features using equation (3), given by,

$$F''_j = FS(F'_j), \quad (3)$$

where $FS(\cdot)$ is a function representing the feature selection algorithm. F''_j is a selected feature vector of j th sample. It uses the feature selection strategy for selecting discriminative features from the combined feature vector of deep Convolutional features of multiple layers and hand-crafted features for scene representation. lines 10 and 11 are representing the training process of the classifier, **Train_Cl**, and lines 12 to 14 show the testing process in which the

trained classifier, Train_Cl , is used to predict the label L_j for j th image. Label indicates the scene type such as SkyGnd is a class label of sky ground image in Fig 1 (see in red color). Finally, line 15 of algorithm 1 shows the performance of the proposed method by measuring the accuracy (Acc.), Precision (Pr), Recall (Re), and F-score (F-Sc) values.

Algorithm for traditional features extracted from the input image I.

Algorithm 2: Traditional_Features_extraction(I)

Input: $I_{w,h,c}$: Input image with 'w' width and 'h' height, 'c' number of channels.

Setting of the Parameters:

mxm: Number of patches

Output: F^{Trad} feature vector

```

1:  $M = \frac{w}{m}$ ,  $N = \frac{h}{m}$ , T=M+N// size of image patch I.
2: For k= 1 to m // rows, e.g. m=4.
3:   For l = 1 to m // columns
4:      $patch_{kl}^{M,N} = I(k \times M + 1 : (k + 1) \times M, l \times N + 1 : (l + 1) \times N, :)$  //size of each patch
// feature for each patch
5:      $[\alpha_{kl}^x, \beta_{kl}^x] = texture\_gradient(patch_{kl}^{M,N})$  // 2 features calculation in x direction.
6:      $[\alpha_{kl}^y, \beta_{kl}^y] = texture\_gradient(patch_{kl}^{M,N})$  // 2 features calculation in y direction.
7:      $[Hog_{kl}^{Bins}] = Hog\_features(patch_{kl}^{M,N}, Bins, \theta)$ ; //Bins=9 and  $\theta = [180^\circ, 360^\circ]$ .
8:      $[R_{kl}, G_{kl}, B_{kl}] = grayWorld(patch_{kl}^{M,N})$  //applying gray world color coefficient algorithm.
9:      $R_{kl} = \frac{1}{T} \sum_{u=1,v=1}^{M,N} R_{kl}^{u,v}$ ,  $G_{kl} = \frac{1}{M+N} \sum_{u=1,v=1}^{M,N} G_{kl}^{u,v}$ ,  $B_{kl} = \frac{1}{M+N} \sum_{u=1,v=1}^{M,N} B_{kl}^{u,v}$  //average the value
//Calculate HSV value
10:     $Sat_{kl} = \frac{1}{T} \sum_{u=1,v=1}^{M,N} (sat_{kl}^{u,v})$ , //taking the average of Saturation (S) component of the color.
11:     $Var_{kl} = \frac{1}{T} \sum_{u=1,v=1}^{M,N} (var(sat_{kl}^{u,v}))$ , // taking the average of variance (V) of Saturation component
12:     $Hue_{kl} = \frac{1}{T} \sum_{u=1,v=1}^{M,N} (Hue_{kl}^{u,v})$ , // taking the average of Hue (H) component
13:     $V_{kl} = \frac{1}{T} \sum_{u=1,v=1}^{M,N} (Val_{kl}^{u,v})$ ,
14:  End for // rows
15: End for // columns
16: Normalized the hog, colors values using equation (1).
17:  $F^{Trad} = \cup_{k=1,l=1}^{m,m} \{ \alpha_{kl}^x, \beta_{kl}^x, \alpha_{kl}^y, \beta_{kl}^y, Hog_{kl}^{Bins}, R_{kl}, G_{kl}, B_{kl}, Sat_{kl}, Var_{kl}, Hue_{kl}, V_{kl} \}$  //concatenation
Return  $F^{Trad}$ 
End algorithm
```

4. Implementation

This section describes the implementation setting of the proposed CNN-HF-FS method. It is coded in MATLAB '2019b' and runs on a Asus Laptop with 32 GB of RAM, a 2.6 GHz i7 processor, and a Nvidia GeForce GTX 1060 with 6GB of RAM. 12-Scene dataset [6] is used in this study, which has 12000 samples in total, where each category consists of 1000 scene images. This dataset consists of indoor and outdoor scene classes with a fixed size of 256X256x3 pixels. The dataset is split into two parts: 80% for training and 20% for testing images. The performance of this method is measured by using four types of performance measures [4, 9, 10]: Acc, Pr, Re, and F-Sc values. In most research, the scene recognition performance is measured by using an accuracy parameter. Thus, Accu is calculated along Pr, Re, and F-Sc. Next, the Multi-layer deep CNN features are extracted by pre-trained ResNet-50 CNN model. It contains multiple consecutive layers. The ResNet is pre-trained on the ImagNet Dataset for further transfer learning purposes. To implement the CNN-HF-FS method using ResNet model, the tenth, twelfth, fourteenth, and sixteenth residual blocks and FC layers of ResNet are used to extract the deep features [10]. Therefore, k =1 is substituted by 10th, k=2 is substituted by 12th, and so on in Algorithm 1. This is because extracting initial layer features increases computational cost and provides detailed information but less semantic. *Relu* the activation function is used in which five different feature vectors have 256,256,512,512, and 1000 dimensions, respectively, and these all

features are combined into a single dimension vector. Next, the handcrafted features are obtained using algorithm 2 by setting $m=4$ and then texture gradient (four features), HOG (9 features, Bines=9), and color (7) Features [10] are extracted from each patch that generates 320 feature vectors for the whole input image. Then, the deep features and traditional features are fused using equation (2) which yields 2,856 dimensions fused vector. Next, the feature selection Techniques are used for selecting the rich features of the image scene (see line 9 of Algorithm 1). The main purpose of using the FS strategy is to reduce the redundancy of the features at the global level to improve the accuracy of the system [51, 53]. For this, the following feature selection techniques are analyzed in this manuscript; reliefF [54], A Mutual Information (MI) based approach [55], L2,1-norm regularized discriminative feature selection for unsupervised learning (UDFS) [56], and feature selection via concave minimization (FSV) [57]. ReliefF estimates the quality characteristics based on how well their values classify data sets that are close to each other. It is an iterative, allocated, and supervised-based technique. MI Based technique uses a selection criterion that shares data between the distribution of the values of a particular attribute and its relationship to a specific class. UDFS The most discriminative or dense value subset is chosen from the whole data set in the given phase. It relies on an unstructured learning strategy. FSV is one of the most useful FS techniques in which the process of selection is integrated into the training of an SVM using a linear coding technique, and the SVM generates the most distinguishing characteristics from the given data set.

Table 1: Results of MI and UDFS feature selection techniques on a 12-scene Dataset.

No. of Features (f)	Features Selection Techniques							
	MI				UDFS			
	Acc	Pr	Re	F-Sc	Acc	Pr	Re	F-Sc
2000f	0.9126	0.9132	0.9122	0.9128	0.9317	0.9327	0.9323	0.9322
1500f	0.9146	0.9156	0.9143	0.9148	0.9337	0.9335	0.9333	0.9332
1000f	0.9142	0.9153	0.9143	0.9146	0.9342	0.9339	0.9340	0.9338
800f	0.8596	0.8630	0.8613	0.8615	0.9362	0.9359	0.9361	0.9358
500f	0.8550	0.8546	0.8566	0.8549	0.9196	0.9193	0.9195	0.9193

Finally, the proposed method uses SVM for classification of purpose due to its effectiveness and efficiency for the scene classification tasks. In the execution of the proposed approach, the SVM setup is as follows: The bias (c) is set to one, and the kernel's value is set to 'linear', with a 10-fold rigorous validation. The pre-trained ResNet-50 is fine-tuned using the traditional Backpropagation with gradient descent approach, with batch size set to 10. The regularization parameter is set to 0.1, epochs are set to 20. The learning rate = 0.0003 for each epoch, and the momentum parameters that control learning speed during the training step are set to 0.9. Results of different feature selection techniques on a 12-scene Dataset, the number of selected features, and recognition accuracy, Precision, Recall, and F-score are given in Table 1, Table 2, and Fig 2.

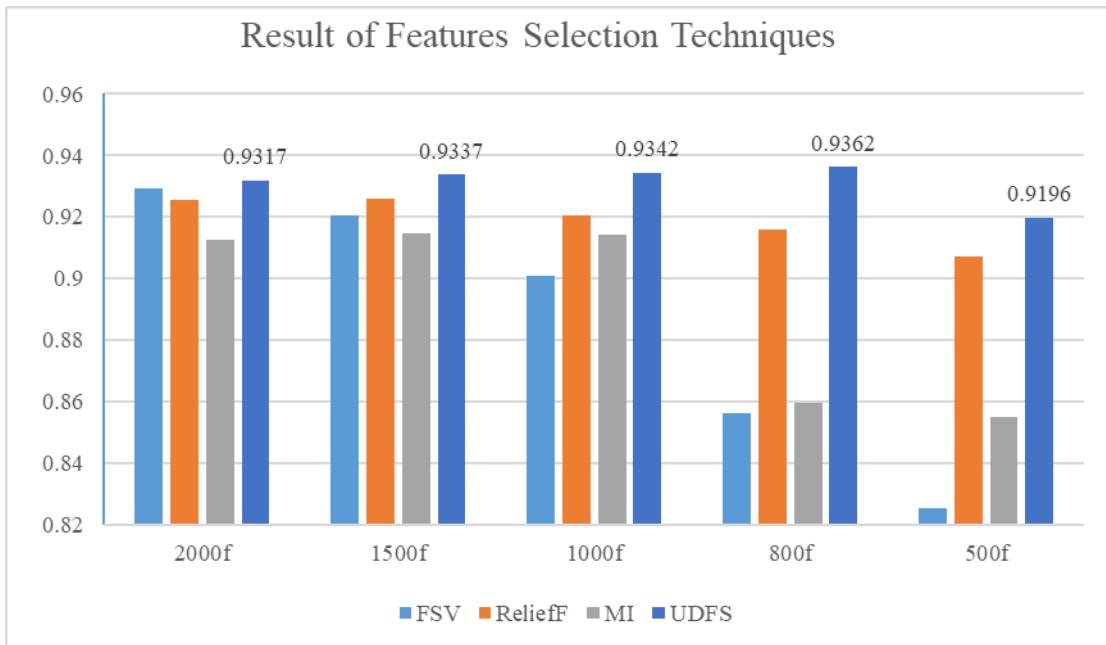


Fig 2: Results of Features selection techniques.

Table 2: Results of FSV and ReliefF feature selection techniques on a 12-scene Dataset.

No. of Features (f)	Features Selection Techniques							
	FSV				ReliefF			
	Acc	Pr	Re	F-Sc	Acc	Pr	Re	F-Sc
2000f	0.9292	0.9282	0.9276	0.9278	0.9257	0.9660	0.9256	0.9256
1500f	0.9204	0.9262	0.9255	0.9255	0.9258	0.9262	0.9255	0.9255
1000f	0.9008	0.8984	0.8988	0.8980	0.9204	0.9212	0.9197	0.9198
800f	0.8562	0.8536	0.8529	0.8532	0.9160	0.9163	0.9158	0.9166
500f	0.8253	0.8237	0.8230	0.8231	0.9071	0.9074	0.9060	0.9064

5. Result and Discussion

5.1 Result of the Proposed Method

The result of the proposed CNN-HF-FS method on 12-Scene dataset using the feature selection technique is shown in Table I and Table II. The accuracy, precision, recall, and F-score performance measured are used in this experiment. To compare the performance of the MI, UDFS (see Table I), and reliefF, FSV (See Table II) feature selection techniques, 2000, 1000, 800, and 500 rich features are selected using them. Then, these features are given to SVM for predicting the class type of input scene. The method without applying FS approaches is also evaluated where it achieves 92.54% recognition accuracy for 20% testing images. The same setting is adopted and then FS approach is involved at the end of the method to select the more discriminative features of the image scene, the results show that the FSV obtained 92.92% recognition accuracy when the 2000 features are selected. However, it reaches 82.53% once the 500 features are selected for scene representation. On the other hand, ReliefF obtains the highest accuracy of 92.97% and lowest accuracy of 90.71% once the 2000 and 500 features are selected, respectively as graphically represented in Fig 2. Next, by applying the MI strategy, it achieves 91.46% highest accuracy when the 1000 features are selected. However, we increase the features up to 2000 but it reduces the accuracy and reaches 91.26%. It is because of reducing the redundancy in the selected features from the features pool. In the last, UDFS feature selection technique is used in which the method reaches 93.62% classification performance and 93.58% F-score when the 800 features are selected for scene representation. It can be observed in Table I that using MI and UDFS techniques, the most discriminative features exist in the range of 800 to 1000, which achieves the highest accuracy of scene representation. Other FS approaches also do not perform well when the number of selected features is increased to more than 2000 or reduced to 500. In this experiment, compared to applying the SVM directly, the FSV, reliefF, and UDFS feature selection techniques show 0.38%, 0.43%, and 1.08%

higher accuracy, respectively, because of selecting the most discriminative features for scene representation. Confusion Matrix of the proposed method using two different feature selection techniques: UDFS and ReliefF are shown in Fig 3, and Fig 4, respectively. As both techniques have a different structure for feature selection, confusion matrices of both FS techniques show that they have variant behavior on similar geometry structures and they show low accuracy when the background objects are added in the images, such as, ‘*personBkg*’, and ‘*Corner*’ classes, because of high similarity among the indoor images.

5.2 Comparison with existing methods and discussion

We compare the suggested method to several advanced scene identification systems. The experimental results are shown in Table III. Applying the proposed CNN-HF-FS method on the 12-scene dataset, it achieves 93.62% classification results, which is higher than other advanced methods [10], [14], [15], by 30.20%, 10.66%, 10.37%, respectively. Meanwhile, the performance of CNN-HF-FS method is compared with the traditional CNN model, [25-28, 49], which shows that it achieves higher accuracy than the standard CNN model as well. The results are given in Table II. On the other hand, compared to advanced approaches e.g., HF-MSF-HFF [16] and Khan et al. [42] methods, the proposed CNN-HF-FS method improves 1.08% and 8.74% recognition accuracy, respectively, when the same setting is adopted for this experiment, see Table I and Table II. This is because the proposed method uses the features selection technique to reduce the redundancy of the features. It also reduces the computational complexity at the classifier level as well.

Accuracy =93.62													
Output Class	skyBkgGnd	96.01	0.00	0.00	0.55	0.46	0.50	0.00	1.40	0.00	0.00	1.09	0.00
	skyGnd	0.00	98.99	0.00	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00	0.55
	bkgGnd	0.00	0.53	99.02	0.00	0.00	0.00	0.00	0.46	0.00	0.00	0.00	0.00
	ground	0.00	0.00	0.00	92.10	0.00	0.94	2.98	0.00	0.00	2.41	1.04	0.54
	sidewalRL	2.04	0.00	0.00	0.00	96.46	0.48	1.01	0.00	0.00	0.00	0.00	0.00
	Box	0.54	0.00	0.00	0.00	0.00	96.38	0.53	0.95	0.53	0.52	0.56	0.00
	diagBKgRL	0.00	0.00	0.00	0.53	2.25	0.97	94.84	0.91	0.00	0.49	0.00	0.00
	groundDiagBkgRL	3.14	0.00	0.94	0.00	0.00	0.99	0.00	88.59	1.54	1.01	3.24	0.56
	Corner	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.90	95.91	0.00	0.53	2.19
	TablePersonBkg	0.00	0.00	0.00	3.29	0.00	0.00	0.00	1.40	0.52	90.95	3.28	0.57
	PersonBkg	0.49	0.00	0.44	0.00	0.00	2.78	0.00	4.33	0.00	6.14	83.18	2.63
	noDepth	0.00	0.52	0.00	1.05	0.00	0.48	0.50	1.79	0.50	0.49	3.15	91.52
Target Class													
skyBkgGnd skyGnd bkgGnd ground sidewalRL Box diagBKgRL groundDiagBkgRL Corner TablePersonBkg PersonBkg noDepth													

Fig 3. Confusion Matrix of CNN-HF-FS when using UDFS technique for feature Selection.

Accuracy = 92.58												
Output Class	skyBkgGnd	98.06	0.00	0.00	0.00	0.98	0.51	0.00	0.46	0.00	0.00	0.00
	skyGnd	0.00	97.54	2.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	bkgGnd	0.00	1.44	97.17	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.49
	ground	0.00	0.00	0.47	93.03	0.00	0.00	0.49	0.00	0.00	4.05	0.52
	sidewalRL	1.01	0.00	0.00	0.00	97.54	0.00	1.00	0.45	0.00	0.00	0.00
	Box	0.51	0.00	0.00	0.00	0.00	94.89	1.51	0.45	0.49	0.00	2.15
	diagBkgRL	0.00	0.00	0.00	0.54	1.93	1.50	94.59	0.90	0.00	0.00	0.54
	groundDiagBkgRL	2.05	0.00	0.00	0.00	0.00	1.52	0.51	85.43	3.50	1.21	3.79
	Corner	0.00	0.51	0.00	0.00	0.00	0.00	0.00	0.48	97.38	0.00	0.57
	TablePersonBkg	0.00	0.00	0.00	1.65	0.00	0.51	0.51	2.30	1.00	89.67	4.36
	PersonBkg	0.47	0.00	0.45	1.00	0.00	1.85	0.92	7.08	1.37	4.95	80.08
	noDepth	0.00	1.78	0.00	1.50	0.00	0.00	0.93	2.51	2.74	1.66	1.99

Target Class

Fig 4. Confusion Matrix of CNN-HF-FS when using reliefF technique for feature Selection.

Table III: Comparison with existing methods. Best results are shown in the Bold.

Method	Accuracy (%)
Khan et al. [10]	63.42
GoogLeNet [28]	82.25
AlexNet [25]	78.13
VGG-16 Net [49]	80.88
ResNet-50 [26]	81.88
Inception ResNet-v2 [27]	84.08
G-MS2F [14]	82.96
FTOTLM [15]	83.25
HF-MSF-HFF [16]	92.54
Khan et al. [42]	84.88
Proposed CNN-HF-FS Method	93.62

The proposed method consists of more semantic features of CNN and handcrafted features, which are linearly combined, and then these features are again preprocessed by FS techniques to obtain the most discriminative features set from the features pool. The proposed method increases the accuracy of the classification of scene type compared to applying the direct classifier at the fused features vector. Furthermore, four different feature selection strategies are applied for extracting 500, 800, 1000, 1500, and 2000 features as results are exhibited in Table I and Table II. In contrast, the computational cost of the proposed technique was not higher than HF-HFF-MSF [16] methods. However, offline FS techniques are applied in addition. Compared to Khan et al. [42] and FTOTLM [15] methods, only offline handcrafted and FS techniques are additionally applied, which shows significant improvement in the accuracy of scene recognition as shown in Table III and Fig 5.

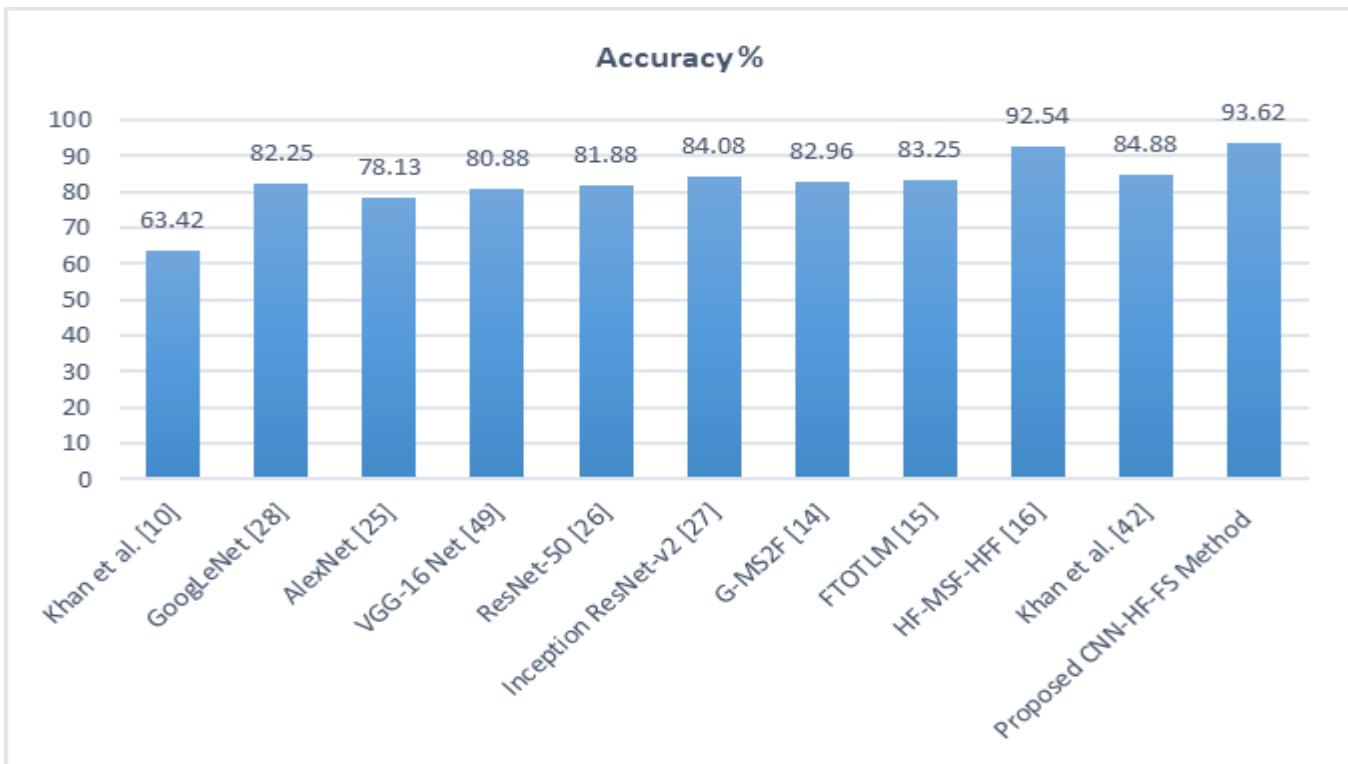


Fig 5.: Accuracy Comparison of the proposed and state-of-the-art methods.

6. Conclusion

In this paper, a robust discriminative feature representation is obtained by applying feature selection techniques to the combined feature vector. The local and global features of pre-trained CNN architecture are mixed with handcrafted features. The feature selection technique is then applied to eliminate redundancy and produce a discriminative feature set for scene representation. Finally, the SVM is combined with a linear kernel to predict the scene type in the input image. The proposed approach surpassed advanced algorithms, attaining 93.62% recognition accuracy on test photos from a 12-scene dataset. Meanwhile, the proposed method did not raise the computational cost of feature extraction when compared to state-of-the-art methods. However, it applies the FS approach offline to reduce feature redundancy. When only a few elements are chosen, the performance of the complicated scene, i.e., the indoor scene, is not remarkable. This is due to a significant loss of background object detail. The proposed method's next work will be to use this discriminative information for object localization and scene segmentation for autonomous cars, merging it with cutting-edge segmentation algorithms.

Declarations

Ethical Approval

Not applicable.

Competing interests

There is no competing of interests.

Funding

No funding was received for this work.

Availability of data and materials

The data will be available on request.

7. References

- [1].Richards, W., A. Jepson, and J. Feldman, *Priors, preferences and categorical percepts*, in *Perception as Bayesian inference*, C.K. David and R. Whitman, Editors. 1996, Cambridge University Press. p. 93-122.
- [2].Anderson, C.H., D.C. Van Essen, and B.A. Olshausen, *CHAPTER 3 - Directed Visual Attention and the Dynamic Control of Information Flow*, in *Neurobiology of Attention*, L. Itti, G. Rees, and J.K. Tsotsos, Editors. 2005, Academic Press: Burlington. p. 11-17.
- [3].Ansari, G.J., et al., *A Non-Blind Deconvolution Semi Pipelined Approach to Understand Text in Blurry Natural Images for Edge Intelligence*. *Information Processing & Management*, 2021. **58**(6): p. 102675.
- [4].Nedovic, V., et al., *Stages as models of scene geometry*. *IEEE Trans Pattern Anal Mach Intell*, 2010. **32**(9): p. 1673-87.
- [5].Oliva, A. and A. Torralba, *Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope*. *International Journal of Computer Vision*, 2001. **42**(3): p. 145-175.
- [6].Khan, A., A. Chefranov, and H. Demirel. *Texture Gradient and Deep Features Fusion-Based Image Scene Geometry Identification System Using Extreme Learning Machine*. in *2020 3rd International Conference of Intelligent Robotic and Control Engineering (IRCE)*. 2020. University of Oxford, UK.
- [7].Lazebnik, S., C. Schmid, and J. Ponce. *Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*. in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. 2006.
- [8].Yang, Y. and S. Newsam, *Bag-of-visual-words and spatial extensions for land-use classification*, in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2010, ACM: San Jose, California. p. 270-279.
- [9].Lou, Z., T. Gevers, and N. Hu, *Extracting 3D Layout From a Single Image Using Global Image Structures*. *IEEE Transactions on Image Processing*, 2015. **24**(10): p. 3098-3108.
- [10]. Khan, A., A. Chefranov, and H. Demirel, *Image-Level Structure Recognition Using Image Features, Templates, and Ensemble of Classifiers*. *Symmetry*, 2020. **12**(7): p. 1072.
- [11]. Sanchez, J., et al., *Image Classification with the Fisher Vector: Theory and Practice*. *Int. J. Comput. Vision*, 2013. **105**(3): p. 222-245.
- [12]. Cheng, X., et al., *Scene recognition with objectness*. *Pattern Recognition*, 2018. **74**: p. 474-487.
- [13]. Zou, J., et al., *Scene classification using local and global features with collaborative representation fusion*. *Information Sciences*, 2016. **348**: p. 209-226.
- [14]. Tang, P., H. Wang, and S. Kwong, *G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition*. *Neurocomputing*, 2017. **225**: p. 188-197.
- [15]. Liu, S., G. Tian, and Y. Xu, *A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter*. *Neurocomputing*, 2019. **338**: p. 191-206.
- [16]. Khan, A., A. Chefranov, and H. Demirel, *Image scene geometry recognition using low-level features fusion at multi-layer deep CNN*. *Neurocomputing*, 2021. **440**: p. 111-126.
- [17]. Zafar, B., et al., *Image classification by addition of spatial information based on histograms of orthogonal vectors*. *PLOS ONE*, 2018. **13**(6): p. e0198175.
- [18]. Ali, N., et al., *A Hybrid Geometric Spatial Image Representation for scene classification*. *PLOS ONE*, 2018. **13**(9): p. e0203339.
- [19]. Giveki, D., *Scale-space multi-view bag of words for scene categorization*. *Multimedia Tools and Applications*, 2021. **80**(1): p. 1223-1245.
- [20]. Meng, X., Z. Wang, and L. Wu, *Building global image features for scene recognition*. *Pattern Recognition*, 2012. **45**(1): p. 373-380.
- [21]. Yuan, L., et al., *Improve scene classification by using feature and kernel combination*. *Neurocomputing*, 2015. **170**: p. 213-220.
- [22]. Ghalyan, I.F.J., *Estimation of ergodicity limits of bag-of-words modeling for guaranteed stochastic convergence*. *Pattern Recognition*, 2020. **99**: p. 107094.
- [23]. Zhou, L., Z. Zhou, and D. Hu, *Scene classification using a multi-resolution bag-of-features model*. *Pattern Recogn.*, 2013. **46**(1): p. 424-433.
- [24]. Lin, G., et al., *Visual feature coding based on heterogeneous structure fusion for image classification*. *Inf. Fusion*, 2017. **36**(C): p. 275–283.
- [25]. Alex, K., I. Sutskever, and G.E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*. 2012: p. 1097–1105.
- [26]. He, K., et al. *Deep Residual Learning for Image Recognition*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [27]. Szegedy, C., et al., *Inception-v4, inception-ResNet and the impact of residual connections on learning*, in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 2017, AAAI Press: San Francisco, California, USA. p. 4278–4284.
- [28]. Szegedy, C., et al., *Going deeper with convolutions*. 2015. p. 1-9.
- [29]. Deng, J., et al. *ImageNet: A large-scale hierarchical image database*. in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009.
- [30]. Hussain, N., et al. *Intelligent Deep Learning and Improved Whale Optimization Algorithm Based Framework for Object Recognition*. 2021.

- [31]. Özyurt, F., E. Sert, and D. Avci, *An expert system for brain tumor detection: Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine*. Medical Hypotheses, 2020. **134**: p. 109433.
- [32]. Khan, M.A., et al., *A resource conscious human action recognition framework using 26-layered deep convolutional neural network*. Multimedia Tools and Applications, 2021. **80**(28): p. 35827-35849.
- [33]. Kwon, Y.-H., S.-B. Shin, and S.-D. Kim, *Electroencephalography Based Fusion Two-Dimensional (2D)-Convolution Neural Networks (CNN) Model for Emotion Recognition System*. Sensors (Basel, Switzerland), 2018. **18**(5): p. 1383.
- [34]. Khan, S., et al., *Human Action Recognition: A Paradigm of Best Deep Learning Features Selection and Serial Based Extended Fusion*. Sensors (Basel), 2021. **21**(23).
- [35]. Liu, S. and W. Deng. *Very deep convolutional neural network based image classification using small training sample size*. in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). 2015.
- [36]. Zhou, B., et al., *Places: A 10 Million Image Database for Scene Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018. **40**(6): p. 1452-1464.
- [37]. Azhar, I.S.M.R.M.K.M.A.U.Y.H.-S.A. *Decision Support System for Face Sketch Synthesis Using Deep Learning Artificial, Intelligence*. Sensors, 2021. **21**, DOI: 10.3390/s21248178.
- [38]. Saleem, F., et al., *Human Gait Recognition: A Single Stream Optimal Deep Learning Features Fusion*. Sensors (Basel), 2021. **21**(22).
- [39]. Wang, C., G. Peng, and B. De Baets, *Deep feature fusion through adaptive discriminative metric learning for scene recognition*. Information Fusion, 2020. **63**: p. 1-12.
- [40]. Liu, B., et al. *Learning a Representative and Discriminative Part Model with Deep Convolutional Features for Scene Recognition*. in Computer Vision -- ACCV 2014. 2015. Cham: Springer International Publishing.
- [41]. Wang, C., G. Peng, and W. Lin, *Robust local metric learning via least square regression regularization for scene recognition*. Neurocomputing, 2021. **423**: p. 179-189.
- [42]. A Khan, A.C., H Demirel, *Building discriminative features of scene recognition using multi-stages of inception-ResNet-v2*. Applied Intelligence, 2023. **53**: p. 18431-18449.
- [43]. Hoiem, D., A.A. Efros, and M. Hebert, *Recovering Surface Layout from an Image*. International Journal of Computer Vision, 2007. **75**(1): p. 151-172.
- [44]. Geusebroek, J.-M. and A.W.M. Smeulders, *A Six-Stimulus Theory for Stochastic Texture*. International Journal of Computer Vision, 2005. **62**(1): p. 7-16.
- [45]. Geusebroek, J.-M., A.W.M. Smeulders, and J. van de Weijer. *Fast Anisotropic Gauss Filtering*. in Computer Vision — ECCV 2002. 2002. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [46]. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005.
- [47]. Xiao, J., et al. *SUN database: Large-scale scene recognition from abbey to zoo*. in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. San Francisco, CA, USA IEEE.
- [48]. LeCun, Y., et al., *Backpropagation Applied to Handwritten Zip Code Recognition*. Neural Computation, 1989. **1**(4): p. 541-551.
- [49]. Simonyan, K. and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv:1409.1556 2015: <http://arxiv.org/abs/1409.1556>.
- [50]. Szegedy, C., et al. *Going deeper with convolutions*. in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
- [51]. Khan, A., et al., *White blood cell type identification using multi-layer convolutional features with an extreme-learning machine*. Biomedical Signal Processing and Control, 2021. **69**: p. 102932.
- [52]. Liang, G., et al., *Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification*. IEEE Access, 2018. **6**: p. 36188-36197.
- [53]. Özyurt, F., *A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine*. Soft Computing, 2020. **24**(11): p. 8163-8172.
- [54]. Liu, H. and H. Motoda, *Computational methods of feature selection*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Taylor & Francis, New York. 2007.
- [55]. Zaffalon, M. and M. Hutter, *Robust feature selection by mutual information distributions*, in Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence. 2002, Morgan Kaufmann Publishers Inc.: Alberta, Canada. p. 577–584.
- [56]. Yang, Y., et al., *L₂-norm regularized discriminative feature selection for unsupervised learning*. 2011. 1589-1594.
- [57]. Bradley, P.S. and O.L. Mangasarian, *Feature Selection via Concave Minimization and Support Vector Machines*, in Proceedings of the Fifteenth International Conference on Machine Learning. 1998, Morgan Kaufmann Publishers Inc. p. 82–90.



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:

Jamil et al. (2024). "From Frames to Events (FFE): Capturing and Merging Video Captions for Temporal Analysis". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024
jamil et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

of time and the continuity of actions. Unlike image captioning, video captioning for multiple video frames is not a good representation of temporal events and interconnection between them. The caption of a single frame might describe objects in the given frame, however, for multiple frames captioning cannot entail the temporal context and interplay of events. One of the key approaches to representing temporal events is to construct graphs that reflect the interconnection between objects as well as how these interactions change over time [3]. In such cases, each vertex represents a Noun while the edges represent verbs as well as propositions. For the above example, "the person" would be a Noun and property of a vertex which will be connected to the second vertex having "Chair", while both are interconnected with the

From Frames to Events (FFE): Capturing and Merging Video Captions for Temporal Analysis

Azhar Jamil ^{a*}

Saif-ur-Rehaman ^b

Hassan Nazeer Chaudhry ^a Farzana Kulsoom ^c

^a Department of Computer Science, Barani Institute of Information Technology, Rawalpindi, Pakistan.

(azhar@biit.edu.pk, hassan@biit.edu.pk).

^b Department of Computer Science, University Institute of Information Technology Rawalpindi, Pakistan (saiif@uiit.edu.pk).

^c Department of Telecommunication Engineering, University of Engineering and Technology, Taxila, Pakistan (farzana.kulsoom@uettaxila.edu.pk).

* Corresponding author: azhar@biit.edu.pk

Abstract:

Video captioning represents a significant advancement beyond static image descriptions by capturing temporal dynamics inherent in visual sequences. This study introduces a novel approach to constructing temporal event graphs from video captions, focusing on the dynamic relationships between entities and their temporal evolution. Each video frame generates captions that undergo preprocessing to identify entities and actions, forming the basis for constructing a graph where entities act as nodes and actions as edges. This paper details a robust methodology for graph construction, utilizing techniques such as part-of-speech tagging and sequential analysis to capture temporal dependencies effectively. The performance of the generated captions was evaluated using ROUGE-N, ROUGE-L, BLEU, METEOR, and CIDEr, with respective scores of 0.485, 0.684, 19.49, 0.642, and 1.658, demonstrating the accuracy and effectiveness of the proposed method. This research underscores the utility of temporal event graphs in advancing applications such as activity recognition and event prediction within knowledge-base systems.

Keywords: NLP; Deep Learning; Scene Captioning; knowledge graphs; Sequence2Sequence; Attention; Word Embedding.

1. Introduction

Image captioning generates a textual description of the image, which describes the visual content of the image encompassing the objects, actions and other relevant details [1]. Another similar captioning task involves videos which is an extension of image captioning since frames are reminiscences of images. Even though video captioning has a resemblance to image captioning, there is a major difference in video captioning. Videos represent events which span over larger window of time also known as temporal events. This means multiple frames of a single window have associated events which progress over a certain window of time [2]. The video captioning encapsulates the dynamic content of a video, capturing both the visual and temporal aspects of the scene.

This means that video frames are not isolated entities, they are interconnected and together constitute temporal events. Each frame in a video represents a scene, and the sequence of frames reveals the progression of actions and interactions [3]. For example, a person sitting on a chair with a cup of tea on the table. Each frame depicts subtle changes: a hand reaching for the cup, lifting it, taking a sip, and then placing it backdown. The sequence of frames portrays the person's interaction with the tea, illustrating the flow

of time and the continuity of actions. Unlike image captioning, video captioning for multiple video frames is not a good representation of temporal events and interconnection between them. The caption of a single frame might describe objects in the given frame, however, for multiple frames captioning cannot entail the temporal context and interplay of events. One of the key approaches to representing temporal events is to construct graphs that reflect the interconnection between objects as well as how these interactions change over time [3]. In such cases, each vertex represents a Noun while the edges represent verbs as well as propositions. For the above example, "the person" would be a Noun and property of a vertex which will be connected to the second vertex having "Chair", while both are interconnected with the

Jamil et al. "From Frames to Events (FFE): Capturing and Merging Video Captions for Temporal Analysis" edge having value "sitting on" [4]. There are several use cases which can be exploited from temporal graphs for example human activity recognition [5], smart city scenarios [6] [7], smart agriculture [8] and various other application domains. There are frameworks which support temporal graph evaluation and pattern detection [9] [10]. Other applications include surveillance, self-driving cars [11], and simultaneous localization and mapping (SLAM) in robotics [12], where understanding the progression of events is critical. The creation of a temporal event graph is further integrated into a knowledge base system that improves our ability to predict future events, answer context-specific questions, and apply logical reasoning to infer hidden relationships. Building a knowledge base system from video scenes is crucial because it allows for a structured representation of dynamic events unfolding over time. This representation not only captures the temporal order of actions and interactions within a scene but also enables deeper analysis of behavioral patterns, event detection, and predictive modelling [4].

1.1. Research Objectives

This study presents several important contributions that advance the field. These include new perspective on the core problem, and a thorough evaluation that improves both the accuracy and applicability of the proposed approach:

1.1.1. Research Objective 1

Utilize existing methods to generate detailed captions for video frames, serving as the foundation for constructing temporal graphs.

1.1.2. Research Objective 2

Implement part-of-speech tagging to accurately identify and classify objects, actions, and their interconnections within video captions, forming the basis for graph construction.

1.1.3. Research Objective 3

Design and validate a novel approach to constructing temporal graphs from video captions, capturing the temporal relationships and context between objects.

1.1.4. Research Objective 4

Utilize the constructed temporal event graphs as knowledge bases to enhance the system's capabilities in event prediction, context-specific question answering, and logical reasoning.

1.2. Research Contributions

This research work presents a novel technique to build temporal graphs describing the interconnection between objects and the temporal evolution of this relationship. As a first step, we take frames from the video and generate a caption from each frame. The second step part of speech tagging (POS) is performed which identifies objects (nouns), actions (verbs) and connections between objects (prepositions). As a next step, a graph is constructed using POS tags generated in step two. To summarize, this study makes the following key contributions:

- a) Extract frames from the video and generate detailed captions for each frame, forming the basis for further analysis.
- b) Utilizes POS tagging to identify objects (nouns), actions (verbs), and their interconnections (prepositions) in the captions, facilitating accurate graph construction.
- c) Introduces a new method to construct temporal graphs from video captions, effectively capturing the temporal context and interconnections between objects.
- d) Leverages temporal event graphs to serve as knowledge bases, enhancing capabilities in event prediction, context-specific question answering, and logical reasoning.

The remaining paper is divided into a related work in section II, the research problem is given in section III, proposed solution in section IV, Experimentation is section V, results and discussion in section VI and the last section VII provides conclusion and future directions.

2. Related Work

Initially, video captioning techniques relied on template-based language models [13]-[15]. These methods follow a bottom-up approach by initially predicting semantic elements like objects, scenes, and activities, and then generating sentences based on language templates that are predefined. These techniques rely heavily on predefined templates and the accuracy of the predicted video concepts, which can limit the diversity of the generated captions. More recently, inspired by advancements in deep learning and neural machine translation (NMT) [16], several sequence learning-based models [17]-[19] have been introduced to tackle the video captioning problem. Treating video captioning akin to a “translation” process, these methods employ encoder-decoder architectures to generate captions extracted from video scenes. Venugopalan et al. [20] pioneered video description using an encoder-decoder architecture, though their approach employed mean pooling across frame features, overlooking temporal sequence information. Subsequent studies [17], [21], [22] addressed this limitation by integrating temporal attention mechanisms and LSTM-based encoders to model long-term temporal structures. Yang et al. [20] improved upon these methods by dynamically identifying regions-of-interest within video frames to extract discriminative features, enhancing the quality of video captions. Xu et al. [23] introduced SeqVLAD, leveraging frame-level feature aggregation to capture detailed spatial information within video content. However, existing methods primarily emphasize global frames or salient regions without adequately distinguishing individual object instances, thus failing to effectively capture the temporal evolution of individual objects within videos. In this study, we introduce an approach (FFE), which constructs temporal graphs across video frames to capture the trajectories of objects over time. Additionally, we employ text preprocessing to filter out irrelevant words from contributing to graph construction. By applying POS tagging, we identify candidate tags, enriching the graph’s role as a knowledge base for tasks such as next event prediction and reasoning.

3. The Research Problem

The research problem centers on the need to effectively represent and analyze the complex temporal relationships between objects within video sequences. In many traditional approaches, video analysis often focuses on individual frames or short segments, neglecting the evolving interactions between objects over time. This results in a significant gap in understanding the full context of the video, which is crucial for applications like event prediction, context-aware question answering, and logical reasoning. One of the main challenges is that current methods struggle to accurately capture and model the dynamic nature of object interactions as they progress throughout a video. For instance, understanding how objects interact and change their relationships over time is essential for predicting future events or making informed decisions based on video content. However, without a robust framework to represent these temporal dynamics, the analysis remains incomplete and less effective. This research addresses this critical issue by proposing a novel technique to build temporal graphs from video captions. These graphs aim to encapsulate the temporal evolution of relationships between objects, providing a more comprehensive understanding of the video content. By generating detailed captions for each frame and using part-of-speech tagging to identify objects, actions, and their interconnections, the research intends to create a structured representation that captures both the spatial and temporal dimensions of video data. The ultimate goal is to leverage these temporal graphs as knowledge bases that enhance the system’s ability to predict future events, answer context-specific questions, and engage in logical reasoning based on the temporal context of the video. This research problem is particularly relevant for advancing the fields of video analysis, artificial intelligence, and machine learning, as it seeks to bridge the gap between static frame analysis and dynamic video understanding.

4. Proposed Solution

This section discusses the proposed technique for temporal graph construction from the video captions. Initially, the frames are extracted from the video as can be seen in Fig. 1. For this paper, we have taken one frame per 5 seconds. Once the frames are extracted the captions are generated for each frame, more details can be seen from subsection IV-A. As the first step, the text captions generated from each frame are prepossessed, further detail is provided in sub-section IV-B. Once the text is preprocessed a graph is constructed defining the temporal sequence of events discussed in sub-section IV-C. The last sub-section IV-C1 defines the overall algorithm and how a graph based on a temporal sequence of events can be used as a knowledge base.

A. Video frame captioning

This section unfolds the methodology of generating descriptive captions from video frames. Fig. 1 shows the overall methodology, the videos are decomposed into individual frames where the frames are represented as $F = \{F_1, F_2, \dots, F_n\}$ extracted at one frame per 5 seconds within a maximum video length of one minute. Captions for

Jamil et al. "From Frames to Events (FFE): Capturing and Merging Video Captions for Temporal Analysis" these frames are generated using the work of [25], the code for this is publicly available. Charades dataset is being used for video captioning, which comprises 9,848 annotated videos depicting daily indoor activities [24]. It is important to note that our research work does not involve caption generation and we mostly focused on temporal graph generation based on existing caption generation work. This module encodes visual features and leverages temporal and spatial contexts to produce coherent natural language descriptions. The generated captions are subsequently passed to the preprocessing module, discussed in Section III-B responsible for eliminating unwanted words and generating POS tags. These POS tags are then used for graph generation as discussed in section III-C.

B. Text Preprocessing

The preprocessing step is crucial in preparing the text captions for analysis. It involves several tasks to clean and standardize the text data, ensuring that it is ready for the subsequent steps of graph construction. As the first step, the text is tokenized, followed by stop word removal. Stop words are common words that do not add significant meaning to the text and can be considered redundant for analysis. Examples include {"the", "is", "in"}. By removing these words, we focus on the more meaningful words that contribute to the construction of the graph. For instance, the phrase "*The man*" becomes "*man*", streamlining the text and emphasizing the essential components. As the second step lemmatization is applied. This process involves converting different forms of verb into the respective infinitive form of the verb. For example, the words "eat", "ate" and "eaten" are all transformed into "eat". This standardization is important because it enables the treatment of numerous forms of the verb as a single entity or action, avoiding unnecessary complexity in the graph. Lemmatization ensures that

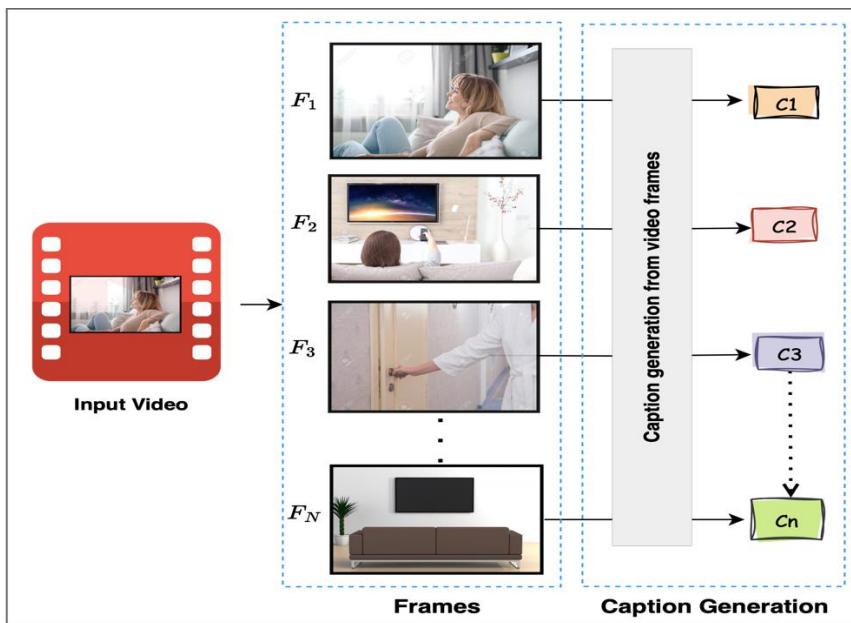


Figure 1: Graph representing event caption from video

variations of a word do not create multiple nodes in the graph, which could lead to fragmented and less coherent representations. As the third step, POS tagging is performed, it identifies the POS segments within the text, such as nouns, verbs, adjectives, and prepositions. The graph is built in such a way that nouns constitute the vertices, while edges represent the verbs and prepositions. This step is critical because it helps distinguish verbs, nouns and prepositions in the sentences which are used to build vertices and edges in the graph. By combining stop word removal, lemmatization, and POS tagging, we ensure that the text data is clean, standardized, and well-structured. This comprehensive preprocessing step is foundational for accurately constructing a temporal graph that effectively captures the sequence of events described in the video captions. It can be seen that the captions generated in Fig. 1, are used as tokens, X_1, X_2 up till X_6 in Fig. 2. It is assumed in the Fig. 2 that all X_i tokens are lemmatized and stop words have been removed from them. It can be seen that once the tokens are passed through POS, the POS tags are obtained which are later used to construct the event graph as discussed further in subsection IV-C.

C. Graph Construction

Graph construction involves transforming text into a structured representation by identifying entities (nouns) as vertices and relationships or actions (verbs and prepositions) as edges. This method captures interactions and dependencies between entities, aiding in tasks like knowledge extraction and scene understanding. The process includes recognizing key entities and determining their relationships through actions or spatial/temporal prepositions, resulting in a graph that reflects the sentence's meaning. For example, the sentence "*the girl is sitting on the couch watching TV*" having nouns such as "girl", "couch," and "TV" become vertices of a graph while relationships like "sitting" and "watching" form the

Jamil et al. "From Frames to Events (FFE): Capturing and Merging Video Captions for Temporal Analysis" edges. This results in a graph with vertices for "girl", "couch" and "TV" and edges connecting "girl" to "couch" labelled "sitting" and "girl" to "TV" labelled "watching", effectively representing the sentence's structure as shown in Fig. 2. The final step is to arrange these nodes and edges to reflect the chronological order of events. By doing so, we create a temporal graph that not only captures the entities and actions but also their progression over time. This graph is invaluable for understanding the flow of events and serves as a comprehensive knowledge base for applications like semantic scene analysis, querying, and analysis.

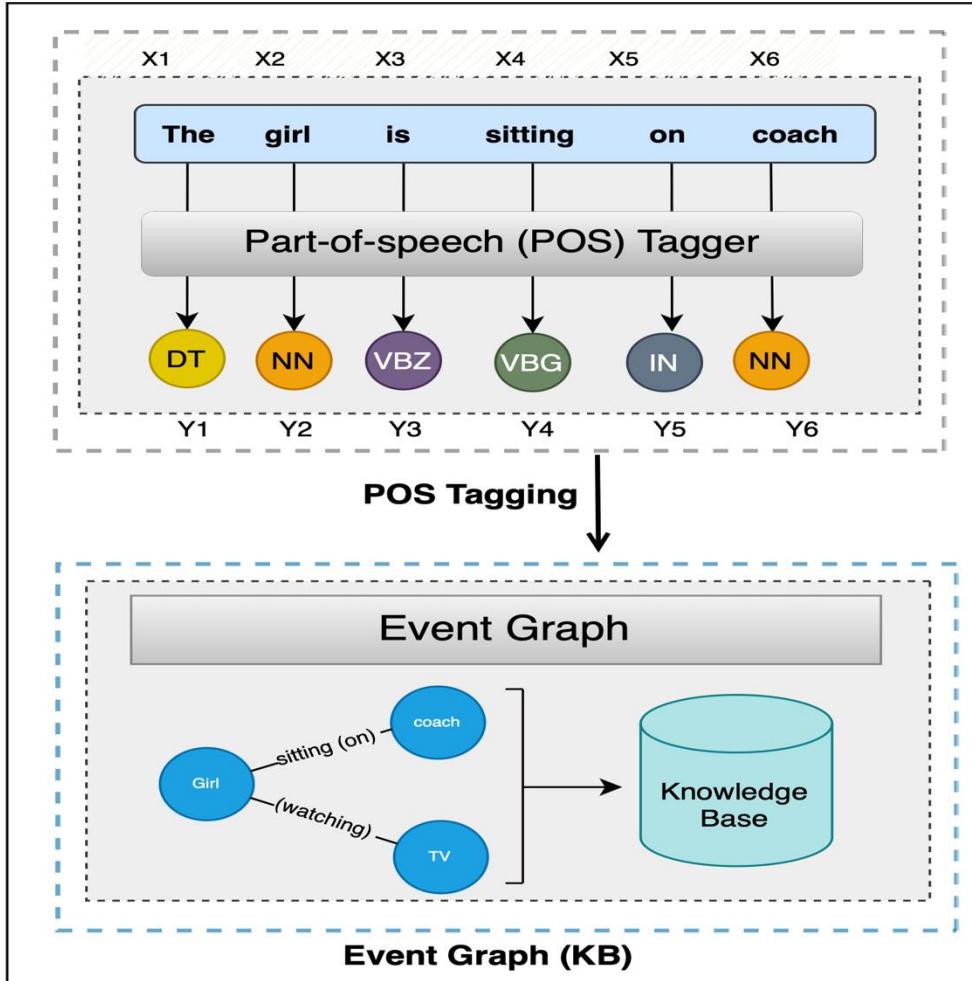


Figure 2: Temporal Event Graph

1) Graph as a temporal sequence of events:

Once the nodes are established, the next step is to determine the temporal relationships between these nodes. This involves arranging the nodes in the order they appear in the video, effectively creating a timeline of events. For instance, if a video shows "a person walking a dog and then the dog sitting", the graph would first have nodes for the "person" and the "dog" linked to the action of "walking", followed by the dog linked to the action of "sitting". This temporal arrangement helps capture the sequence in which events occur, providing a coherent flow of activities and interactions. This temporal sequencing is vital for applications that require an understanding of the progression of events, such as activity recognition and event prediction.

D. Graph Algorithm

Proposed algorithm for constructing the temporal graph from video captions is detailed here. Starting with an empty graph ' G ', each caption from the list of video frames is being processed and then tokenized the text to break it into individual words or tokens to perform POS tagging that identify their grammatical roles. From these tagged tokens, we extract the entities and actions, which form the core of graph. Each entity and action are added as a node to graph G , with entities labelled as '**entity**' and actions as '**action**'. This clear differentiation helps define the nature of each node. Once all nodes are added, we identify pairs of related entities and actions, representing their interactions or relationships like a person acting or an object being acted upon. For each identified pair, we add an edge to the graph G , labelling it with the type of relationship identified. This process continues until all captions are processed and all relevant nodes and edges are added to the graph. The final graph G then provides a structured, comprehensive representation of the

Algorithm 1: Temporal Graph Construction from Video Captions

Require: List of captions from video frames

Ensure: Graph G representing the knowledge base

```

1: Initialize an empty graph G
2: for each caption c in captions do
3:   tokens  $\leftarrow$  tokenize(c)
4:   pos tags  $\leftarrow$  part of speech tagging (tokens)
5:   entities actions  $\leftarrow$  extract entities actions (pos tags)
6:   for each entity action in entities actions do
7:     if entity action is an entity, then
8:       add node (entity action, type = 'entity') to G
9:     else
10:      add node (entity action, type = 'action') to G
11:    end if
12:  end for
13:  for each pair ( $e_i, a_j$ ) in get related pairs (entities actions) do
14:    add edge ( $e_i, a_j$ , relation= identify relation( $e_i, a_j$  )) to G
15:  end for
16: end for
17: return G

```

E. Misplaced Relationship (MR)

MR is a phenomenon where labels of vertices and edges are correct, however, the link between them is incorrect. For example, if a man sits on a chair and watches the TV, in the perfect world the graph should be made like the Man (vertex) is connected to a chair (vertex) with the relationship (sits on) between them. The second relationship should be the Man (vertex) is connected to the TV (vertex) with the relationship (watches). In MR the chair (vertex) gets linked up with the TV (vertex) with a relationship (watches). MR can be due to the following reasons:

- 1) *Ambiguity in Language:* Natural language often contains ambiguities that can confuse algorithms. For example, in the sentence "*The man sits on the chair and watches TV*", the algorithm might misinterpret which verb applies to which noun due to the proximity and structure of the sentence. The word "watches" might mistakenly be linked to "chair" instead of "man."
- 2) *Contextual Misunderstanding:* Algorithms might not fully understand the context. While humans can use common sense to understand that a chair cannot watch TV, an algorithm might not have this contextual awareness if it relies purely on syntactic patterns without semantic understanding.
- 3) *POS Tagging Errors:* POS tagging might incorrectly label the words in a sentence. For instance, if the algorithm incorrectly tags "chair" as an entity that can perform actions, it might erroneously link it with "watches."

5. Experiments

We have used the publicly available Charades dataset [24], which consists of 9,848 videos of daily indoor activities. The average length of videos is 30 seconds, and it contain free text descriptions at the video level. For our research work, captions were required at frame level. We have chosen 254 videos randomly for the Charades dataset and chose 1 frame per 5 seconds, generating 5 captions per video. The original data set has 46 objects and 30 verbs however, the chosen subset of videos is reduced to 34 objects and 54 action classes.

6. Results and Discussion

For this experiment, we have compared hand-annotated graphs with the generated graphs in terms of part of speech. The MSE is calculated between POS given in hand-annotated graphs and graphs generated by proposed algorithms. The results are shown in Fig. 3a. It could be seen that prepositions remain mostly in place increasing the number of

Jamil et al. "From Frames to Events (FFE): Capturing and Merging Video Captions for Temporal Analysis" prepositions does not change much. In almost all cases accuracy MSE remains below 0.20 which is quite reasonable. It is important to note that the graph for prepositions could only be tested till 6 for sentences such as, "*The cat jumped over the fence and ran through the yard before hiding under the car near the garage*". The remaining values given in Fig. 3a are extrapolated and are not generated by our algorithm. For verbs and nouns, the error remains reasonable till 6, after this point increasing the number of verbs or nouns drastically increases the error. For edges (verbs) the error increases very rapidly and climbs up to 0.9 for 10 verbs. For example, "*After he wakes up, he brushes his teeth, eats breakfast, reads the newspaper, jogs in the park, takes a shower, gets dressed, drives to work, attends meetings, and writes emails*". The generated graphs have 90% incorrect captions of edges if compared to the ground truth. For nouns, the acceptable error is till the number of nouns is 6, although the MSE of nouns is less than for verbs and at 8 nouns it becomes stable, however for 10 nouns half of the vertex labels are misplaced as compared to ground truth.

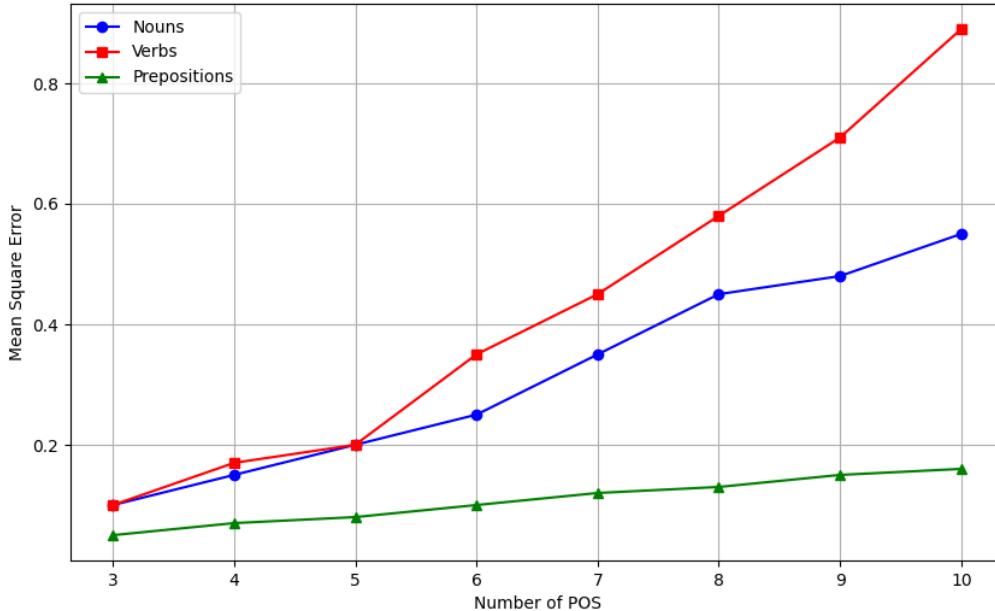


Figure 3 (a): POS mean square error

A. MSE of MR

To assess the total number of MR we have increased the size of the graph in terms of vertices and edges from 3 to 10. We compared hand-annotated graphs with generated graphs, Fig. 3b shows the result. It could be observed that the graph size of five edges and vertices the misplaced vertices or edges are one while is 10% of error. After 6 there is a total of 4 incorrect edges and 2 incorrect edges which makes a total of 6 incorrect graph elements out of 12 resulting in 50% of error. It could be inferred that the proposed technique works well for graph sizes less than equal to 5 vertices and edges. It is important to highlight that increasing the number of edges (connections) results in more errors as compared to increasing the number of vertices.

B. Captioning accuracy

We measured the accuracy of the generated captions against a set of annotated ground-truth captions. The following table summarizes the accuracy results for a sample of videos from the dataset. We have used two ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores, the Rouge-N measures the overlap of N-Gram between the generated and annotated captions while Rouge-L compute the longest common subsequence between generated and reference captions. The third column in the table is Perplexity which determines how well a probability model predicts a sample. BLEU (Bilingual Evaluation Understudy), evaluates the precision of N-Grams between generated and annotated captions. METEOR (Metric for Evaluation of Translation with Explicit ORdering), takes precision, recall, word order, stemming and synonymy into account. Finally, the most relevant criteria to our work, CIDEr (Consensus-based Image Description Evaluation), computes the similarity of a generated caption to multiple reference captions, it also takes into account the consensus among human annotators. It employs a Term Frequency-Inverse Document Frequency (TF-IDF) weighting scheme to assign higher weights to important words and phrases.

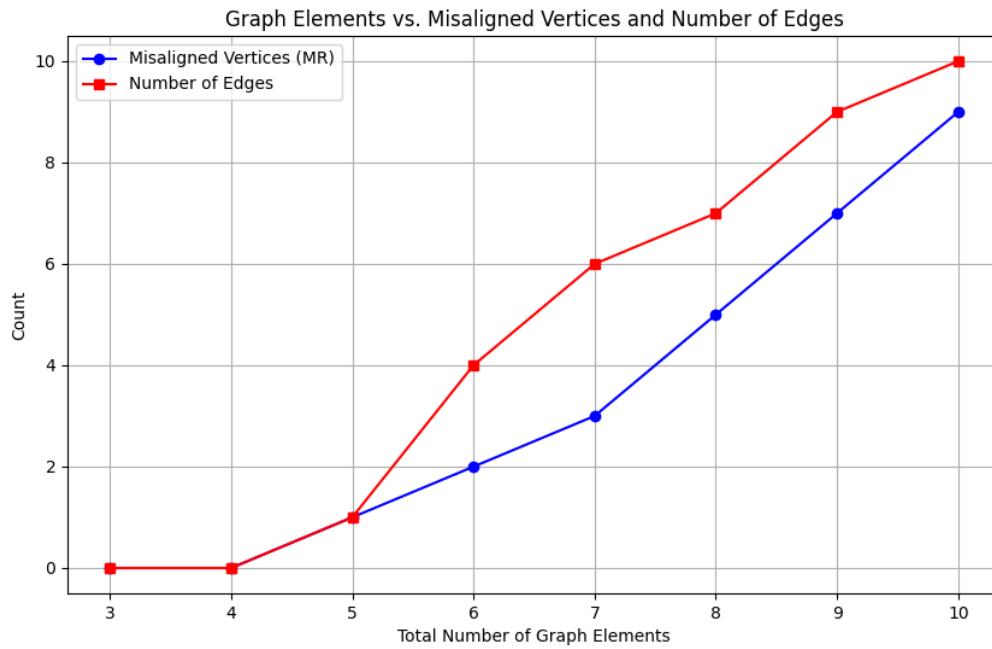


Figure 3 (b): Graph elements versus MR

Table 1 shows the results for each frame, we have taken the average in the last row indicated in bold. Rouge-N (0.478) means that the generated captions for Frame 1 overlap approximately 47.8% with the annotated ground-truth captions in terms of N-Gram overlap. Similarly, for Rouge-L (0.685), The overlap between the generated and reference captions for Frame 1 is approximately 68.5%, indicating a strong match in sequence similarity. Perplexity (20.25) indicates that the language model used for generating captions has a perplexity of 20.25 on Frame 1, suggesting moderate predictive capability. A BLEU score of 0.65 shows that the precision of N-Grams in Frame 1's generated captions against the reference is about 65%, indicating fairly accurate phrase matching. METEOR (0.56) takes precision, recall, word order, stemming, and synonymy into account; Frame 1 achieves a METEOR score of 0.56, reflecting moderate overall quality compared to the reference. The last column, CIDEr (1.702), evaluates similarity to multiple references, indicating a reasonable consensus among annotators on the quality of generated descriptions. A CIDEr score of 1.702 indicates a moderate level of agreement or the degree of similarity observed between the captions generated and the various reference captions. In conclusion, for all given metrics the generated captions are reasonable and could be used to build the knowledge base. Generally, Perplexity ranges between 18.43 and 21.05 is acceptable, in our experiment, it is 19.49. Similarly, BLEU is between 0.63 to 0.67 suggesting a moderate to good precision in capturing N-Gram matches between generated and reference captions, our BLEU score is 0.642. For METEOR the overall quality considering various linguistic factors ranges between 0.55 to 0.58, we achieved 0.564. For CIDEr the acceptable range is between 1.600 to 1.702, in our case, it is 1.658.

Table 1: Comparison of Evaluation Metrics for Generated Captions

Frames No.	Rouge-N	Rouge-L	Perplexity	BLEU	METEOR	CIDEr
Frame 1	0.478	0.685	20.25	0.65	0.56	1.702
Frame 2	0.512	0.700	18.43	0.67	0.58	1.680
Frame 3	0.495	0.690	19.70	0.64	0.57	1.665
Frame 4	0.460	0.670	21.05	0.62	0.55	1.600
Frame 5	0.482	0.675	20.00	0.63	0.56	1.645
Total	0.485	0.684	19.49	0.642	0.564	1.658

7. Conclusions and Future Direction

In conclusion, this paper has presented a novel methodology for constructing temporal event graphs from video captions, emphasizing the interconnectedness of entities and their temporal evolution. By leveraging techniques like part-of-speech tagging and sequential analysis, we have demonstrated the effectiveness of our approach in capturing complex temporal dependencies. Evaluation using metrics such as ROUGE-N, ROUGE-L, BLEU, METEOR, and CIDEr has validated the accuracy and utility of the generated captions, highlighting their potential in applications such as activity recognition and event prediction.

As future work, the integration of graph neural networks (GNNs) holds promising avenues for advancing our research. GNNs can enhance the representation learning within temporal event graphs, enabling more sophisticated reasoning about temporal dynamics and entity interactions. Future work could explore the application of GNNs for refining entity embeddings, predicting future states based on historical event graphs, and improving the robustness of temporal event understanding in dynamic visual content. Moreover, extending our methodology to handle real-time video streams and large-scale datasets could further broaden the applicability and scalability of our approach in real-world scenarios.

Acknowledgment

We are thankful to the research reviewers who helped in writing and improving these guidelines.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

8. References

- [1] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in international conference on machine learning. PMLR, 2015, pp. 2048–2057.
- [2] J. Zhang and Y. Peng, "Video captioning with object-aware spatio-temporal correlation and aggregation," IEEE Transactions on Image Processing, vol. 29, pp. 6209–6222, 2020.
- [3] B. Pan, H. Cai, D.-A. Huang, K.-H. Lee, A. Gaidon, E. Adeli, and J. C. Niebles, "Spatio-temporal graph for video captioning with knowledge distillation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10 870–10 879.
- [4] S. Li, Z.-F. Zhang, Y. Ji, Y. Li, and C.-P. Liu, "Spatio-temporal graph-based semantic compositional network for video captioning," in 2022 International Joint Conference on Neural Networks (IJCNN). IEEE, 2022, pp. 1–8.
- [5] F. Kulsoom, S. Narejo, Z. Mehmood, H. N. Chaudhry, A. Butt, and A. K. Bashir, "A review of machine learning-based human activity recognition for diverse applications," Neural Computing and Applications, vol. 34,no. 21, pp. 18 289–18 324, 2022.
- [6] S. Awan, Z. Mehmood, H. N. Chaudhry, U. Tariq, A. Rehman, T. Saba, and M. Rashid, "Profiling casualty severity levels of road accident using weighted majority voting." Computers, Materials & Continua, vol. 71, no. 3, 2022.
- [7] J. Liu, X. Wang, H. Lin, and F. Yu, "Gsaa: A novel graph spatiotemporal attention algorithm for smart city traffic prediction," ACM Transactions on Sensor Networks, 2023.
- [8] M. San Emeterio de la Parte, S. Lana Serrano, M. Muriel Elduayen, and J.-F. Martínez-Ortega, "Spatio-temporal semantic data model for precision agriculture iot networks," Agriculture, vol. 13, no. 2, p. 360, 2023.
- [9] H. N. Chaudhry and M. Rossi, "Optimising queries for pattern detection over large scale temporally evolving graphs," IEEE Access, 2024.
- [10] P. Daverio, H. N. Chaudhry, A. Margara, and M. Rossi, "Temporal pattern recognition in graph data structures," in 2021 IEEE International conference on big data (Big Data). IEEE, 2021, pp. 2753–2763.
- [11] L. Ye, Z. Wang, X. Chen, J. Wang, K. Wu, and K. Lu, "Gsan: Graph self-attention network for learning spatial–temporal interaction representation in autonomous driving," IEEE Internet of Things Journal, vol. 9, no. 12, pp. 9190–9204, 2021.
- [12] A. Walcott-Bryant, M. Kaess, H. Johannsson, and J. J. Leonard, "Dynamic pose graph slam: Long-term mapping in low dynamic environments," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2012, pp. 1871–1878.
- [13] M. Rohrbach, W. Qiu, I. Titov, S. Thater, M. Pinkal, and B. Schiele, "Translating video content to natural language descriptions," in ICCV, 2013, pp. 433–440.
- [14] N. Krishnamoorthy, G. Malkarnenkar, R. J. Mooney, K. Saenko, and S. Guadarrama, "Generating natural-language video descriptions using text-mined knowledge," in AAAI, 2013, pp. 541–547.
- [15] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in ICLR, 2015, pp. 1–15.

[16] D. Bahdanau, "Neural machine translation by jointly learning to align and translate," in ICLR, 2015, pp. 1–15.

[17] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko, "Sequence to sequence-video to text," in ICCV, 2015, pp. 4534–4542.

[18] J. Wang, W. Wang, Y. Huang, L. Wang, and T. Tan, "M3: Multimodal memory modelling for video captioning," in CVPR, 2018, pp. 7512–7520.

[19] J. Zhang and Y. Peng, "Hierarchical vision-language alignment for video captioning," in MMM, 2019, pp. 42–54.

[20] S. Venugopalan, H. Xu, J. Donahue, M. Rohrbach, R. Mooney, and K. Saenko, "Translating videos to natural language using deep recurrent neural networks," in ACL, 2015, pp. 1494–1504.

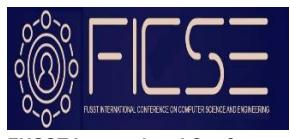
[21] L. Yao, A. Torabi, K. Cho, N. Ballas, C. Pal, H. Larochelle, and A. Courville, "Describing videos by exploiting temporal structure," in ICCV, 2015, pp. 4507–4515.

[22] P. Pan, Z. Xu, Y. Yang, F. Wu, and Y. Zhuang, "Hierarchical recurrent neural encoder for video representation with application to captioning," in CVPR, 2016, pp. 1029–1038.

[23] Z. Yang, Y. Han, and Z. Wang, "Catching the temporal regions-of-interest for video captioning," in ACM MM, 2017, pp. 146–153.

[24] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, "Hollywood in homes: Crowdsourcing data collection for activity understanding," 2016. [Online]. Available: <https://paperswithcode.com/dataset/charades>

[25] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6077–6086.



FUSST International Conference
on Computer Science and
Engineering

FICSE

Research Article

Article Citation:

Sultan et al. (2024). "Instrument to investigate value-based thinking in the study of Software Engineering". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024 Sultan et al.



Published by
Foundation
University Islamabad.
Web: <https://fui.edu.pk/>

because software and software development are so ubiquitous these days (e.g., software supports or enables any product, and innovations through software frequently drive changes to societies and communities). As a result, the concept of value may be much broader and examined in terms of the value of the software package or service, but it also has connections to the practices, processes, artifacts, and principles of software development, as well as to the development organizations, teams, and individual developers themselves. In addition, as software engineers, we would

Instrument to Investigate Value-Based Thinking in the Study of Software Engineering

Muhammad Umar Sultan^a Dr. Amber Sarwar^b

^a Department: Computer Science Rawalpindi Women University
(umar.sultan@f.rwu.edu.pk)

^b Department: Computer Science Rawalpindi Women University
(amber.sarwar@f.rwu.edu.pk).

Abstract:

A significant threat in software industry is that of project failure. Due to the inability of the developers to understand "value-focused" elements and develop alternatives in the project, the project cannot be completed within the set time and resources which leads to project failure. Value-Added thinking is one of the methods that are supposed to help to concentrate the decision-making process on those activities that are to be performed in order to solve the problem. According to the thinking of value, the community needs to have this thinking in order to have a successful software development which can be put in the education of software engineering. This also entails the assessment of options at the architectural level. They reduce the failure probability of the project by channeling the thinking that is strategic in the project. To solve these issues. It is necessary to explain and strengthen the students' ability to comprehend the value-focused thinking. They help in assessing the value-based thinking in teaching of software engineering courses..

Keywords: Value-based thinking, software engineering, instruments, pilot test.

1. Introduction

Values are central to decision-making in value-focused thinking (VFT). Explicit values facilitate communication and negotiation, the generation of new alternatives, the classification of current options, and the discovery of fresh avenues for decision-making. When something occurs and you have to determine how to react, that choices present a difficulty; when you intentionally decided to step back, that decision presents an opportunity [1].

Any activity that generates no value for the user or customer is considered waste. Value is defined as whatever the customer is willing to pay for. Software engineering's primary goal is to create and manage high-caliber software-intensive systems and services that benefit both users and society at large. In addition, software engineering endeavors must generate the most feasible benefit to aid developers in accomplishing their assignments within the limitations of time, money, and resources. For instance, developers might "value" technologies that assist them in finding the smallest possible collection of test cases, but they might view lengthy planning sessions as a "waste" because they don't aid in producing a high-caliber product [2].

Like lean development, any activity that neither consumers nor developers believe has feature value is considered waste and presents risks (e.g., project completion delays, dissatisfied customers and developers). However, in software engineering, explicit value and waste issues have traditionally received little attention. To be more precise, while availability, security, safety, and pricing have all drawn some attention, software engineering pays little attention to larger human values (including the idea of import beyond corporate, economic, and technological worth) like compassion, social responsibility, and justice [2].

Furthermore, value features in software and its engineering are fetching even more crucial because software and software development are so ubiquitous these days (e.g., software supports or enables any product, and innovations through software frequently drive changes to societies and communities). As a result, the concept of value may be much broader and examined in terms of the value of the software package or service, but it also has connections to the practices, processes, artifacts, and principles of software development, as well as to the development organizations, teams, and individual developers themselves. In addition, as software engineers, we would

like to address values that go beyond those of the software development communal, such as automation, production, superiority, low cost, etc. [2].

Since value-focused thinking emphasizes on the numerous pronouncement objectives and how many alternatives are generated from decision rewards, and the CEO seeks to determine unfamiliar solutions to the subject, it may be a more creative approach. There is a claim that this departure time facilitates achieving the necessary outcomes of the decision. When employing value-focused reasoning, one strives to achieve the most straightforward possible result by starting there [4].

To achieve economic leverage, a novel concept is applied during the development of Value-Focused Software (VFS). When an innovative idea is introduced, there is a high degree of uncertainty regarding the outcome. There is an extreme amount of ambiguity in the design of these systems. In the production phase of those systems, it is very desirable that important stakeholders be involved in achieving desired results [5].

The effectiveness of Value Focused Software Engineering (VBSE) processes greatly affects the productivity of VFS software development. The software engineering (SE) doctrines that are in use now and are evolving are combined with the value principles in the value-focused approach, and these doctrines all work together to create a new paradigm [6].

A definition of VBSE provided by Barry Boehm and colleagues is "the clear consideration of value issues in the application of science and mathematics that results in the qualities of computer software that are beneficial to humans" [7]. Realizing an original idea is the primary question in VBSE. A unique proposal has never been acknowledged by the industry before, and it is doubtful that its viability will be determined prior to implementation [8]. But over time, "the study of value's non-economic and monetary components was added to the purview of VBSE" [7]. Researchers still define value in terms of an object's economic or monetary worth.

A humanistic technique called value-based thinking (VBT) is used to help in decision making. Its three main concepts are as follows: begin with values, utilize values to generate the best alternatives, then evaluate those options using values.

- Values come first. Start with the goals of the stakeholders and decision makers rather than with potential alternatives.
- Provide the best substitutes. Once values are understood, apply them to create superior substitutes.
- Consider values when assessing options. Lastly, use the values to the multiple verifiable decision analysis operations-research approach to assess alternatives.

It is a team sport, VFT. To produce higher-value options for potential clients, the entire team of analysts and decision-makers have to understand the humanities and work together actively. Keeney lists nine advantages of value-based reasoning. Three of these advantages are especially relevant to operational analysis, even if the majority are evident. The first is directing strategic thought: the commander's goals for actions can be captured by value-focused thought. The second is assessing options; different courses of action can be assessed through multiple objective decision analysis. The third step involves developing alternatives. After evaluating them, we can determine value gaps—the discrepancy between the ideal value and the best alternative—and concentrate our efforts on coming up with better ones [18].

A decision-making process called the value-focused thinking approach lays out a procedure for determining values and methodically arranging them. When users are making decisions, these objectives are important to them. make up the value portfolio. It lays forth the requirements for each choice from the decisionmakers' points of view. An objective means-ends network of fundamental goals and means goals can be connected to by the VFT technique [20]. Values are the things that matter to those who make decisions. A stakeholder's ability to rank, prioritize, balance, and trade-off priorities can have a significant impact on their reasoning and evaluation of decommissioning results and decisions. The values of stakeholders may be influenced by a variety of things, such as earlier experiences, religious convictions, and personal interests. developed the decision science theory known as value-focused reasoning, which builds on the notion of varying stakeholder values [21].

Even though critical thinking skills are highly valued, there is still considerable debate in the literature about what exactly qualifies as a critical thinking talent, how to develop it, and how to apply it. According to the Delphi Report, which represents the consensus of numerous experts across various disciplines, critical thinking is defined as "purposeful,

self-regulatory judgment that leads to interpretation, analysis, evaluation, and inference, along with the explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations underlying that judgment." Some definitions focus on the ability to examine, analyze, and evaluate mental processes, while others highlight an attitude of inquiry. Critical thinking and introspective problem-solving are both necessary for clinical reasoning in the healthcare sector [25]. As the metaverse grows, it becomes increasingly important to comprehend how to provide and receive value in this new digital environment. Creating engaging, meaningful, and relevant entertainment for consumers is an essential component of adding value in the metaverse. This entertainment can take many different forms, such as virtual interactions, games, and user-generated content including environments, objects, and avatars, among other things [26].

1. Related work

Many research works have been done on software engineering with a value focus. Emily Winter et al. [9][10] conducted two studies on the role of humanoid standards in software engineering. They looked at Advancing the research of Human Values in Software Engineering in the first research [9]. In this paper, the case is made for the importance of studying human values in Software Engineering (SE) as an emerging field of study that will have a big impact on society. It presents two fundamental ideas to further this research agenda: the importance of values as different from ethics, yet connected to it; and the need of well-defined theoretical frameworks for values study. A Values Q-Sort tool created in accordance with these two concepts is used to display emerging data from the first study, which involved 12 participants. Emily, Winter, and Maria Angela Ferrario examined humanoid standards in software engineering in their second study [10]. They talked on how important it is to comprehend how human values work, create instruments and methodologies for studying them in the framework of software engineering, and build on this comprehension to think about how software engineering research could support the development of a more socially conscious software industry. The focus of both research [9][10] has been software engineering's human values. However, the suggested work differs in that it centers software engineering education around value-focused thinking.

Likewise, research on value Paul Ralph and associates engage in focused software engineering [11]. They outlined goals and clarified the many wastes that arise during the software development process. This study offers the first waste-related taxonomy and is empirical in nature. This lists nine wastes, discusses their root causes, underlying conflicts, and general relationship to the Lean Software Development waste taxonomy. However, the generalization of statistics is not accepted by this grounded theory-focused study. Although the proposed taxonomy appears to be broadly applicable, unlike software development values inside firms will lead to different kinds of unused.

Three studies on value-based software engineering were carried out by Barry Boehm et al. [12][13][6]. A method for moving toward Value- Focused Software Engineering is provided in the first study [12], along with its benefits on a national or international scale [14]. It concentrated on its activities, aids, and results at the joint software and information technology (SW/IT) level, in keeping with the soul of concurrent software and system engineering. in the spirit of system and software engineering in parallel. The ultimate goal is to acquire and apply foundational information. They introduced the "4+ 1" Value-Focused Software Engineering (VBSE) hypothesis in the second research [13]. The win-win Stakeholder Theory W, which poses the question, " Which values are important?" is the central engine. "And what can be guaranteed to succeed? "For the aforementioned software engineering firm. Utilities theory is one of the four additional ideas it uses. In third research [6], they employed the value-based approach to software development, which integrates value concerns with both current and developing software engineering ideas and practices while creating a broad framework within which different approaches harmoniously support one another.

Even so, every study mentioned above included a substantial report on the different frame works. A 4+1 theory for value-based software engineering was proposed in one study [13], and the overview and research agenda are represented by other academics [12]. The researcher adopted a value-based approach to develop software in a different frame work [6]. However, the suggested work differs from these in that it approaches software engineering from an educational standpoint while emphasizing value-based thinking.

Similar to that, Anne Am eels et al. [15]. a report regarding VBM, or merit-focused management. This review of the literature examines the value-focused management approaches of six consultants: Stern Stewart & Co., Marakon Associates, McKinsey & Co., Price Waterhouse Coopers, and L.E.K. Consul-ti. This is because VBM is said to be transforming financial management at the highest level in some of the largest companies in the world. The primary topic of this study is the value-focused management control method, which facilitates value creation through integration. To this end, the author has used a sample literature review for research purposes. To the best of our knowledge, no other study has evaluated alternative energy scenarios by combining VFT with MCDM. The primary conclusions of the study are as follows: (i) a consensus objective system for evaluating energy alternatives is created using VFT; (ii) opinions regarding these objectives are gathered from diverse stakeholders using MAUT; and (iii) four different energy transition options in Germany are evaluated. We therefore include those who make decisions at every step of the process [22]. Value is a complex concept that, depending on the topic of study, can be understood and expressed in a variety of ways. Keeney's description of values as guiding moral principles evaluation of options "guidelines for assessing the appeal of any potential substitutes or consequences" is what we use in our inquiry. There exist values at both the individual and communal levels. Values are shared ideals that individuals trade with one another within a community,

organization, or culture. Researchers assert that analyzing the ideas held by members of a group can help to both practically and advantageously explain the general tendencies of collective action [3, 15, 30, 33][23].

Establishing a comprehensive list of objectives is one of the most crucial steps in the first divergent thinking stage of the decision-making process. Early goal-setting in a project will help decision-makers produce clear thinking in both the quantitative and qualitative domains. examination. Value-Focused Thinking (VFT) is a potentially useful technique that can assist in defining objectives and options³¹ in collaboration with key stakeholders. Value-Centered Thought will provide a set of guidelines and objectives for decision-making that direct the creation of more creative alternatives. According to a number of academic studies³², examining a set of objectives can be a useful tool for decision-makers when they need to make choices³³. The decision-makers can achieve the decision's goals by asking the right questions of them [24].

Despite their importance, previous research has noted the value of value-based thinking in software engineering. None of them, meanwhile, concentrate on integrating value-based thinking into software engineering curricula.

There are three open research questions.

RQ1: What is the perceived level of awareness of value-based thinking in software engineering education?

RQ2: What methods contribute to value-based thinking in the teaching of software engineering?

RQ3: How may these techniques be applied in software engineering education?

Research Methodology

A. Survey

The Software Engineering Institute's (SEI) Kasunic [16] guidelines will be followed in conducting the survey. Since his work is the most popular and extensively used guide for carrying out a successful survey in the world of software engineering, we followed it.

2. “Tool to assess Value-Based Thinking in Software Engineering Curricula”

- 1) To what extent are you aware of value-focused thinking?

Foundational Moderate Sufficient Expert

Questions:	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1) Do you believe that evaluating students' awareness of value-focused thinking in software engineering education is necessary?					
2) Do you believe that software engineering education at a university can make use of the value-focused thinking method?					
3) In your opinion, is it necessary to have an understanding of values in order to minimize human error?					
4) Do you believe that all students studying software engineering education need to be aware of value-focused thinking?					
5) Do you believe that teaching students about value-focused thinking should come before incorporating it into software engineering curricula?					

Section:1 Consciousness

Section:2 Practices

Questions:	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
6) Do you believe that teaching software engineering should incorporate guiding strategic thinking?					
7) Do you believe that teaching software engineering students to evaluate alternatives is helpful?					
8) Do you believe that teaching software engineering should include creating alternatives?					
9) Do you believe that developing alternatives, assessing alternatives, and guiding strategic thinking are essential components of a university education?					

10) Which practice is more important in your opinion?

- Providing direction for strategic thought
- Assessing alternatives
- Developing alternative

11) Which practice, in your opinion, can help you reduce waste material while you're creating a system?

- Providing direction for strategic thought
- Assessing alternatives
- Developing alternative

Section:3 How to disseminate practices

Questions:	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
12) Do you believe that incorporating value-focused thinking techniques into software engineering undergraduate courses is a good method to spread these practices throughout the field of software engineering education?					
13) Do you think that requiring the usage of value-based thinking methodologies in the class base project is a good way to propagate the practices in software engineering study?					
14) Do you think that conducting workshops to promote the adoption of value-based thinking methodologies in projects is a useful way to propagate these practices in software engineering study?					
15) Do you believe that requiring value-focused thinking techniques to be applied in the senior project is a means of disseminating the techniques in Education in software engineering?					

16) How might value-focused thinking techniques be more widely applied in software engineering curriculum?

- Participate in the research
- Apply in each project in class
- Lead a workshop
- Apply in the final project

3. Result and Analysis

A. VFT Awareness in Software Engineering Curriculum

This survey part was created to monitor students' perceived understanding of value-based thinking in the context of software engineering study. This segment of the questionnaire consists of two parts. The first portion uses a four-point Likert scale: basic = 0.20, intermediate = 0.15, advance = 0.65, and expert = 1. The second component had a five-point Likert scale with a range of 1 to 5 (Strongly Agree: 5, Agree: 4, Neutral: 3, Disagree: 2, Strongly Disagree: 1). The mean perceived awareness of value-based thinking was estimated in this area. The perceptions of students on their level of expertise and awareness regarding value-based thinking in software engineering education are shown in Figures 1 and 2, respectively.

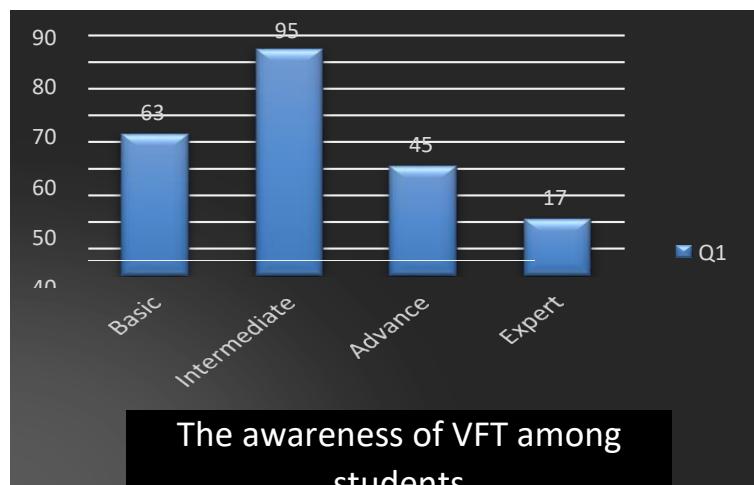
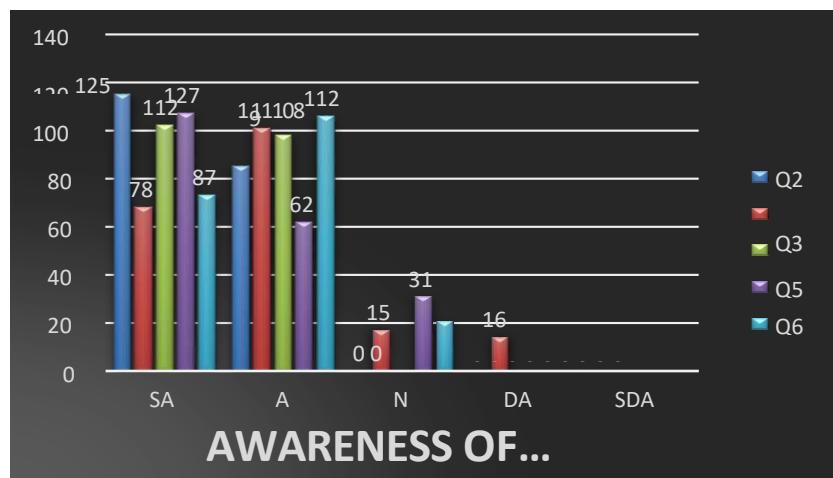


Figure 1. The awareness of VFT among students

The quantity of students according to proficiency level is shown in Figure 1: Awareness of VFT Level. Out of 220 students, the figure shown in this section shows that 63 have a basic understanding of value-focused thinking in



software engineering education, 95 have an intermediate understanding, 45 have an advanced understanding, and 17 have a professional understanding.

Figure: 2 views of students regarding the use of VFT in software engineering courses.

Figure 2 shows the thoughts of students on VFT comprehension in software engineering study. The figure in this section shows five different questions. There are 125 students out of 220 who strongly agree with question number two, 95 students who agree with question number two, none who are neutral toward question number two, and none who disagree or strongly disagree with question number two. According to the Q3 results, 78 students out of 220 strongly agree with Q3, 111 agree with Q3, 15 have no opinion about Q3, and 16 disagree with Q3. Likewise, Q3 is largely agreed with by all 220 students polled. The Q4 figure shows that, out of 220 students, 112 strongly agree with Q4, 108 agree with Q4, 0 are neutral, 0 disagree with Q4, and 0 strongly disagree with Q4. Out of 220 students, 127 strongly agree with Q5, 62 agree with Q5, 31 are neutral toward Q5, 0 disagree with Q5, and 0 strongly disagree with Q5. This

is indicated by the Q5 number. Likewise, the Q6 result displays that 87 out of 220 students strongly agree with Q6, 112 students agree with Q6, 21 students have no opinion about Q6, and zero students strongly disagree with Q6

B. Student Opinions Regarding VFT Methods in Software Engineering Education

The persistence of this survey segment was to check how students felt about the value-focused thinking methods employed in software engineering study. This segment of the questionnaire consists of two parts. In the first part, a five-point Likert scale is employed, with 1 denoting severely disagree, 5 agree, 4 neutral, 2 disagree, and 5 strongly disagree. A three-point Likert scale was provided in the next section; the values were 0.15 for directing strategic thought, 0.65 for assessing options, and 1 for generating alternatives. The mean perception of value-based thinking approaches was estimated in this section. Figures 3 and 4 show, respectively, what students believe to be more useful value-based thinking strategies in software engineering education. These techniques remain more suited for studying software engineering.

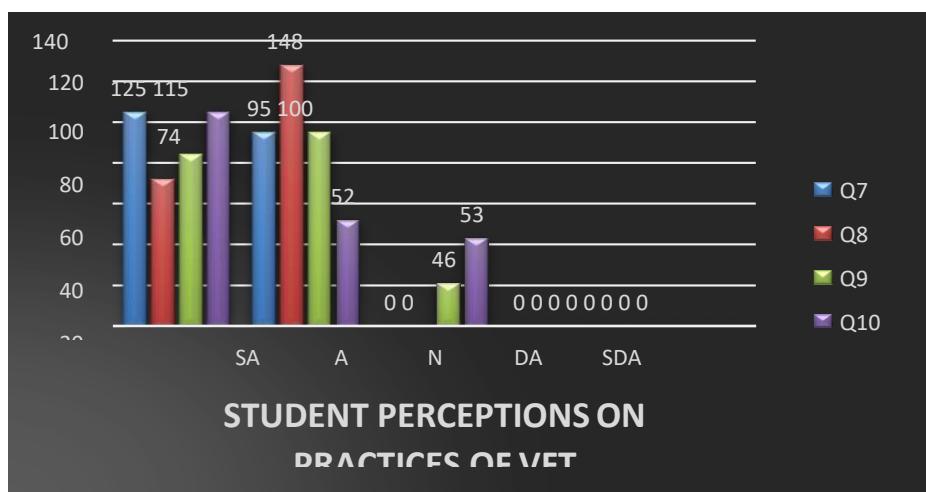


Figure: 3 Students' perceptions of VFT practices

Figure 3 shows how students view VFT processes in software engineering courses. The figure in this section shows four different queries. In reference to Question 7, the data shows that, out of 220 students, 125 strongly agree with the question, 95 agree with the question, 0 are neutral about the question, and 0 disagree with the question and Q7 badly disagree with each other. Out of 220 students, the Q8 figure shows that 72 strongly agree with Q8, 148 agree with Q8, none have a neutral opinion of Q8, none disagree with Q8, and 0 students strongly disagree with Q8. Q8 is strongly disagreed with by 220 students. No student out of 220 disagrees with Q9, and no student out of 220 strongly disagrees with Q9, according to the Q9 number. Comparatively, Q9 is agreed upon by 74 out of 220 students firmly, agreed upon by 100 out of 220 students, and neutrally agreed upon by 46 out of 220 students. The data pertaining to Question 10 shows that, among the 220 students, 115 strongly agree with the question, 52 agree with it, 53 are neutral about it, and none of the 220 students disagree or strongly disagree with it.

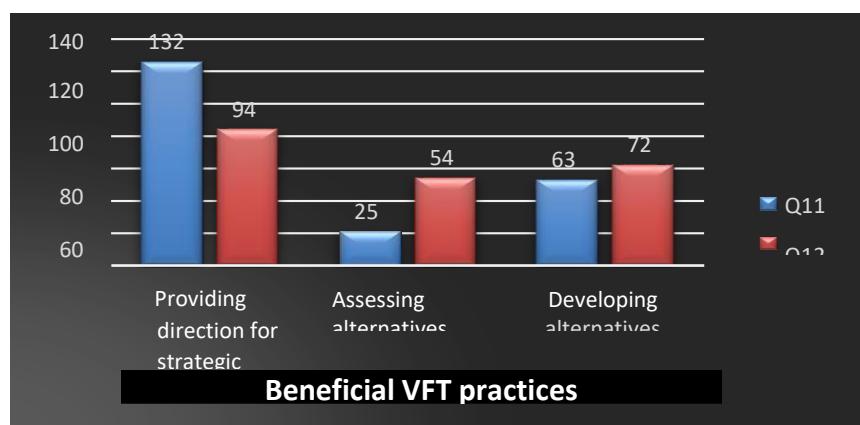


Figure: 4 beneficial VFT practices

Figure 4 compares the value-based thinking methodologies used in software engineering education to see which is more advantageous. Figure 4 poses two questions. Find out which of the eleven practices is more advantageous. Out of the 220 students, 132 think that strategic techniques that direct are more advantageous, 25 think that software engineering education methods that evaluate alternatives are more advantageous, and 63 think that software engineering education practices that create alternatives are more advantageous. A similar issue, "Is procedure better for cutting down on waste when creating a system?" is posed in question 12 out of the 220 students, 94 think that leading strategic initiatives is a more beneficial way to cut waste during development processes, 54 think that assessing initiatives is a more beneficial way to cut waste during development processes, and 72 think that creating alternatives is a more beneficial way to cut waste during development processes.

C. Views on the adoption of VFT techniques in software engineering curriculum.

This section of the study was created to see the opinions of students regarding the dissemination of value-based thinking methods in software engineering instruction. This segment of the questionnaire consists of two parts. In the first portion, a five-point Likert scale is employed, with 1 denoting severely disagree, 5 agree, 4 neutral, 2 disagree, and 5 strongly disagree. Within the second area, there was a four-point Likert scale with values ranging from 0.25 to 1 (Apply in every class project = 0.25, Conduct workshop = 0.65, Apply in final year project = 1). A comprehensive summary of the general consensus regarding the dissemination of value-based thinking approaches in software engineering study can be found in Appendix D. Figures 5 and 6 show the ideas that students have about promoting value-focused thinking processes in software engineering education. Diffusion strategies and the most effective strategy for promoting value-based thinking in software engineering education are other topics on which students have perspective.

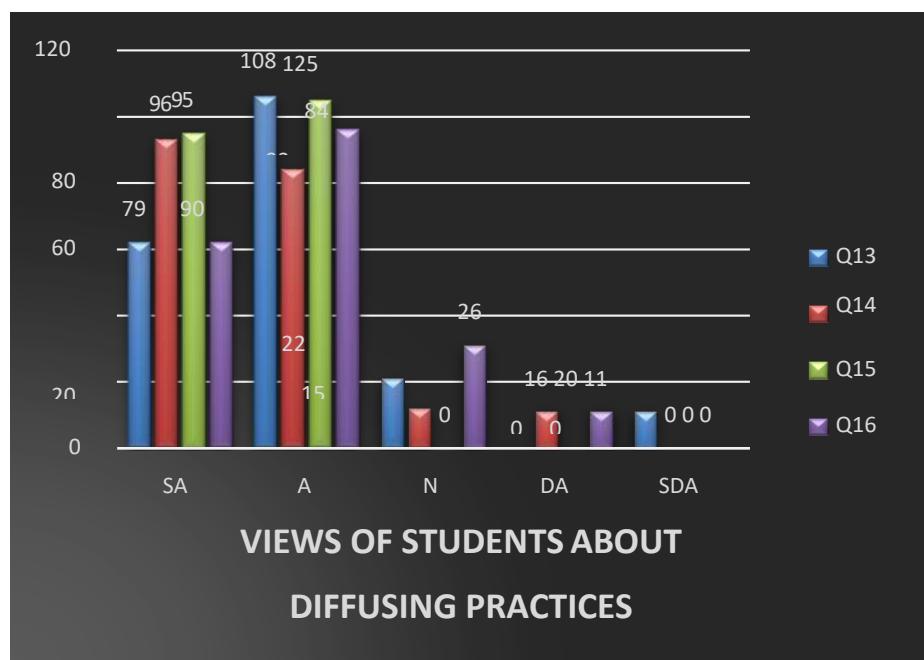


Figure: 5 views of students regarding the adoption of VFT techniques in software engineering curriculum

Figure 5 shows how students think VBT practices should be spread in software engineering study. The figure in this segment shows four different queries. Regarding Question 13, the data shows that 79 out of 220 students strongly agree with the question, 108 students agree with the question, 22 students have no opinion, and no student disagrees. Similarly, 11 out of 220 students strongly disagree with the question. Regarding Q14, the data shows that 96 out of 220 students agree, 93 out of 220 students strongly agree, 15 out of 220 students are neutral, 16 out of 220 students disagree, and 0 out of 220 students strongly disagree with the question.14. Parallel to this, the Q15 number shows that, out of 220 pupils, 95 strongly agree, 125 agree, 0 are unsure, and 0 disagree with the question. Likewise, none of the 220 pupils significantly disagreed with the inquiry. In reference to Question 16, the data shows that 90 out of 220 students strongly agree with the question, 84 students agree with the question, 26 students have no opinion, 20 students disagree with the question, and none of the 220 students strongly disagree with the question.

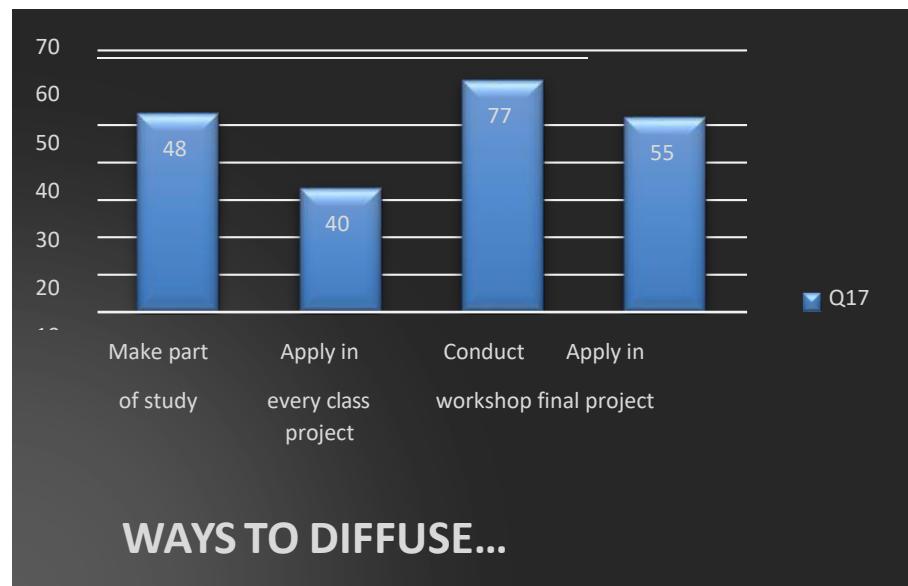


Figure 6 provide strategies for dispersing practices.

Students' ideas about how to propagate value-based thinking techniques in software engineering study are contrasted in Figure 6. The figure in this segment shows that 48 out of 220 students think that value-focused thinking strategies are included in the study, 40 out of 220 say that value-based strategies are used in every class project, 77 out of 220 say that workshops are used to spread value-focused strategies, and 55 out of 220 say that the final project uses value-focused strategies.

4. Limitations

This study's scope will be limited to the following. This study will only focus on value-based thinking in software engineering study. Only academic staff and students at Pakistani universities that grant degrees in software engineering are the study's target population. A student from the fourth to the eighth semester is included in the approximate number of samples for each student. All of the faculty members in the department of software engineering are included in the approximate number of samples for faculty.

5. Future Work

In order to validate the study's findings, a comparison of the survey data from other foreign universities will be conducted with faculty members and students from those institutions who provide software engineering degrees.

6. Conclusion

This study adds three new insights to the field of software engineering study. The first influence of the study will be to the knowledge of the significance of concentrated thought in the education of software engineers. The second contribution will be a list of methods for teaching software engineering with a value-based perspective. The final component of the study will be a technique that may be utilized in software engineering education to spread value-focused thinking. This research will significantly add to the software engineering body of knowledge (SEBOK) by providing a mechanism that may disperse values in the curriculum for software engineering study. This research is expected to help software engineers improve their ability to think value-focus

REFERENCES

- [1] Vaniver, *Value-Focused Thinking: a chapter-by-chapter summary*. 2013.
- [2] Journals.ELSEVIER, "Special Issue on Value and Waste in Software Engineering," 2020. [Online]. Available:[https://www.journals.elsevier.com/inf information-and-software-technology/call-for-papers/special-issue-on-value-and-waste-in-software-engineering](https://www.journals.elsevier.com/inf-information-and-software-technology/call-for-papers/special-issue-on-value-and-waste-in-software-engineering)
- [3] R. L. KEENEY, *Value-Focused Thinking*. 1996.
- [4] S. Tvedt, "Understanding the role of value-focused thinking in idea management Understanding the Role of Value-Focused Thinking in Marcus Selart and Svein Tvedt Johansen," 2011.
- [5] K. El, S. Quintin, and N. H. Madhavji, "User Participation in the Requirements Engineering Process: An Empirical Study," pp. 4–26, 1996.
- [6] A. C. Study and B. Boehm, "Value-Focused Software Engineering: A Case Study," no. March, pp. 33–41, 2003.
- [7] et al, Biffl, S., *Value-Focused Software Engineering*. 2006.
- [8] J. Gordijn and R. J. Wieringa, "AValue-Oriented Approach to E-business," pp. 390–403, 2003.
- [9] E. Winter, S. Forshaw, L. Hunt, and M. A. Ferrario, "Advancing the Study of Human Values in software Engineering," 2019.
- [10] A. E. Winter and M. A. Ferrario, "Studying human values in software engineering," pp. 2019–2021, 2019.
- [11] T. Sedano and P. Ralph, "SoftwareDevelopment Waste," 2017.
- [12] B. Boehm and B. Boehm, "1 Value-Focused Software Engineering: Overview and Agenda."
- [13] B. Boehm and A. Jain, "An Initial Theory of Value- Focused Software Engineering An Initial Theory of value-Focused Software Engineering," no. January 2005, 2014.
- [14] B. W. Boehm, K. J. Sullivan, and T. Hall, "Software Economics: A Roadmap," 2000.
- [15] A. Ameels, P. W. Bruggeman, and A. Ameels, "Value- Focused management control processes to create value through integration a literature review value-focused management: An integratedapproach to value creation literature review.,," 2002.
- [16] M. Kasunic, "Designing an Effective Survey," no. September 2005.
- [17] Drevin, Lynette & Kruger, Hennie & Steyn, Tjaart. (2006). Value-Focused Assessment of Information Communication and Technology Security Awareness in an Academic Environment.
- [18] Chapter 19Value-focused Thinking Gregory S. Parnell United States Military Academy at West Point and Innovative Decisions Inc.
- [19] Henschel, Thomas. (2010). Typology of risk management practices: An empirical investigation into German SMEs. International Journal of Entrepreneurship and Small Business- Int J Enterpren Small Bus. 9. 10.1504/IJESB.2010.031922.
- [20] Gao, Shang, Ying Li, and Hong Guo. "Understanding the value of using smartphones for older adults in China: a value-focused thinking approach." *Digital Transformation for a Sustainable Society in the 21st Century: 18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019, Trondheim, Norway, September 18–20, 2019, Proceedings* 18. Springer International Publishing, 2019.
- [21] Tung, Aaron. "A value focused thinking approach to decommissioning decision making." *SPE Symposium: Decommissioning and Abandonment*. SPE, 2021.
- [22] Höfer, Tim, Ruediger von Nitzsch, and Reinhard Madlener. "Using Value-Focused Thinking and Multi-Criteria Group Decision-Making to Evaluate Energy Transition Alternatives." (2019).
- [23] Soja, Ewa, et al. "Supporting Seniors in Independent and Healthy Ageing through ICT: Insights from a Value-Focused Thinking Study in Latvia, Poland and Sweden." (2021).
- [24] LERA SILVA, CRISTIAN SERGIO. "Envisioning public spaces through value-focused thinking." (2018).
- [25] Denial, Aurora. "The impact of the perceived value of critical-thinking skills." (2024).
- [26] Mirza, Rsha, et al. "Clustering potential metaverse users with the use of a value-based framework: Exploiting perceptions and attitudes on the use and adoption of metaverse for bold propositions." *Telematics and Informatics* 87 (2024): 102074



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:

Zia et al. (2024). "MAED U-Net: A Multi Attentional Encoder Decoder Model for Skin Lesion Segmentation". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

MAED U-Net: A Multi Attentional Encoder Decoder Model for Skin Lesion Segmentation

Usman Zia ^{a*}, Abdur Rehman ^b,

^a Department of Electrical & Computer Engineering, Foundation University, Islamabad, Rawalpindi, Pakistan.
(usman.zia@fui.edu.pk).

^b Department of Electrical & Computer Engineering, Foundation University, Islamabad, Rawalpindi, Pakistan.
(abdur.rehman@fui.edu.pk).

* Corresponding author: usman.zia@fui.edu.pk

Abstract:

The most important and related task in the computer-aided diagnosis of skin cancer is automated skin lesion segmentation. In this paper, a novel deep learning model called MAED U-NET is proposed. Its purpose is to improve the precision and effectiveness of skin lesion segmentation, an essential stage in the automated diagnosis of skin cancer. This model includes dual attention processes to highlight significant features in dermatoscopic images, building upon the strengths of the Residual Neural Network (ResNet) architecture. We assess our model's performance using the HAM10000 and ISIC 2018 dermatological datasets. Using a novel combination of spatial and channel-wise attention inside the ResNet framework, our method allows the model to choose highlight key features and suppress irrelevant data. Our thorough trials show that this leads to enhanced segmentation accuracy. The ISIC 2018 dataset provides a baseline for comparison against other cutting-edge models. The model is further assessed on this dataset after being trained and validated on the HAM10000 dataset, which is renowned for its varied depiction of skin lesions. The findings show that in skin lesion segmentation tasks, the MAED U-NET performs noticeably better than conventional ResNet models. On HAM10000 dataset, MAED U-Net achieves test accuracy of 0.972 accompanied by the DSC score of 0.926 and on ISIC 2018 data set, MAED U-Net achieves test accuracy of 0.958 along with DSC score of 0.842 Our results imply that dual attention methods can be successfully added to deep learning models to improve their feature discernibility in medical image analysis, especially in the difficult area of dermatoscopic image segmentation. In addition to offering a fresh method for segmenting skin lesions, this work establishes the foundation for further research into the use of sophisticated attention mechanisms in medical image processing.

Keywords: Skin Lesion Segmentation, Convolutional Neural Network, Residual Neural Network, Dermoscopy Images.

1. Introduction

One of the most prevalent cancers worldwide, skin cancer poses a serious threat to public health. effective treatment and positive patient outcomes depend on early detection and precise diagnosis. as a crucial diagnostic technique that allows for a thorough inspection of skin lesions, dermatoscopy has gained prominence. on the other hand, dermatoscopic image interpretation is sensitive to inter-observer variability and demands a high level of knowledge [1]. Because of this, there is an increasing interest in creating automated methods to help dermatologists with diagnosis. Medical image analysis has been transformed by deep learning, a subset of artificial intelligence. residual neural networks (ResNet), one of the deep learning architectures, have demonstrated encouraging performance in image identification applications. however, in complicated medical images, such dermatoscopic images of skin diseases, typical ResNet models can have difficulty focusing on the most important features [2].

1.1. Model

The attention-attention ResNet, a cutting-edge deep learning model that enhances the performance of conventional ResNet for skin lesion segmentation, is presented in this thesis. the model improves its focus on relevant characteristics in dermatoscopic images by utilizing both spatial and channel-wise attention processes. the goal of this innovative method is to lessen the shortcomings of current models in terms of identifying small but important characteristics of skin lesions.

1.2. Data Set

Our work uses the important dermatological datasets ISIC 2018 and ham10000 to validate and train the attention-attention ResNet. with its extensive range of skin lesions, the HAM10000 dataset offers a thorough training environment. the model's performance is assessed and contrasted with current state-of-the-art methods using the ISIC 2018 dataset, which is renowned for its benchmarking criteria.

1.3. Contributions

Our primary contributions in this study include:

- Develop Attention-Attention Residual Network Model which can improve the segmentation accuracy of skin cancer.
- Test our model on both these data sets (HAM10000 and ISIC 2018) to perform the cross validation of our model.
- Perform a comparative analysis of the results achieved by our proposed model with results achieved by the other existing deep learning models on HAM10000 and ISIC 2018 data set.

This article examines the attention-attention ResNet model's technological advancements as well as possible applications in dermatology. this model could significantly help physicians with early diagnosis by boosting skin lesion segmentation accuracy, which would improve patient care and treatment outcomes for skin cancer.

2. Literature Review

One of the most prevalent types of cancer in the world, skin cancer requires precise diagnosis and treatment stratification [3]. Because skin lesions can vary greatly in appearance, segmenting skin cancer lesions from dermatoscopic pictures has always been an important but difficult undertaking. Several conventional image processing methods were used to address this problem prior to the development of deep learning technologies [4]. This section examines these conventional techniques, illuminating their workings, advantages, and drawbacks.

2.1 Overview of Traditional Methods

Thresholding: A simple but effective technique in which pixel intensities are divided into foreground (lesion) and background (healthy skin) segments according to a threshold value. One often-used way for automatically figuring out this threshold is Otsu's method [5].

Region-Based Segmentation: Images were segmented using methods such region growth and split-and-merge according to predetermined standards like texture similarity, color, or intensity [6]. Subjectivity was incorporated into the segmentation process by these methods, which frequently required manual selection of seed points or beginning region.

Edge Detection: To determine the boundaries between lesions and healthy skin, edge detection techniques including the Sobel, Canny, and Prewitt operators were applied. These methods rely on the identification of intensity gradients, but they frequently have trouble distinguishing between overlapping lesions [7].

Clustering Techniques: Lesion segmentation is aided by clustering algorithms such as K-means and fuzzy C-means, which group pixels into clusters according to similarities. These techniques occasionally failed with heterogeneous lesion presentations while being unsupervised [8].

2.2 Related Work

Numerous methods for semantic segmentation have surfaced in recent decades, taking advantage of deep learning and machine learning developments. Interestingly, deep learning methods—particularly CNN-based methods—have shown to perform better in automatically segmenting skin lesions [18]. We examine many CNN-based methods for skin lesion segmentation in this section. Yu et al. [19] created the FCRN model, showing that performance peaks at 50 layers and decreases at 101 levels because of insufficient training data. They achieved accuracy of 0.929 to 0.949 on the ISIC 2016 dataset. Yuan et al. [20], [21] introduced the CDNN technique, which by adding a seven-channel input and optimizing with the Jaccard distance loss function, greatly improved accuracy to 0.963 on ISIC 2016 and 0.953 on

ISIC 2017. Artifacts and photos with poor contrast were challenges for the approach. For ISIC 2017, Venkatesh [22] presented a CNN-based architecture with multi-channel RGBH input and residual connections, claiming an accuracy of 0.93. Dash et al. [23] encountered difficulties with lesions resembling normal skin, but they were able to reach an accuracy of 0.948 on a bespoke dataset after modifying the U-Net design to incorporate batch normalization. Tang et al. [24] created SU-SWA, a U-Net-based model that demonstrated accuracies ranging from 0.966 to 0.971 on the PH2 and ISIC datasets. It featured separable convolutional blocks and stochastic weight averaging. Due to its large number of parameters, the model was less appropriate for devices with little resources. Goyal et al. [25] achieved increased segmentation precision with a dice score of 0.892 and Jaccard index of 0.821 using ensemble deep learning approaches (R-CNN and DeeplabV3+). Because these models underwent a great deal of pre- and post-processing, they demanded substantial computer resources. For the ISIC 2016 and 2017 datasets, Lin et al. [26] introduced an encoder-decoder-based architecture that produced accuracies of 0.953 and 0.925, respectively. The requirement for pre- and post-processing stages was removed by this concept. To improve segmentation performance, with notable accuracies on various ISIC datasets, Arora et al. [27], Wibowo et al. [28], and Sarker et al. [29] introduced modifications to the U-Net architecture and other models, incorporating techniques like group normalization, bidirectional ConvLSTM, and kernel factorization. Tong et al. [30] improved feature representation capabilities for effective skin lesion segmentation with accuracies of 0.954 to 0.943 on ISIC and PH2 datasets by incorporating attention gates into the U-Net architecture. Self-calibrated convolutional U-Net and dual encoder architectures are examples of models that Wang et al. [31], Khouloud et al. [32], and Wu et al. [33] proposed. These models achieved high accuracies (up to 0.979) on ISIC datasets, though some models had issues with irregular lesion boundaries or high computational complexity. GFANet was first presented by Qiu et al. [34]. It achieved accuracy of up to 0.970 on several datasets, while segmenting tiny lesions proved to be difficult. With accuracies of 0.918 to 0.885 on ISIC datasets, Zeng et al. [35] and Li et al. [36] investigated structural entropy and uncertainty self-learning strategies in unsupervised learning, highlighting difficulties in low-contrast picture segmentation. A number of researchers proposed modified U-Net models that improved segmentation performance by adding normalization and attention mechanisms. Notable implementations of these models were made by Arora et al., who achieved 0.95 accuracy on ISIC 2018, and Wibowo et al., who included ConvLSTM layers and achieved dice indices of up to 0.951 across multiple datasets. Convolutional neural networks (CNNs) like the U-Net model are made to segment images quickly and accurately. Image segmentation is the process of dividing an image into several segments or pixels that have a common property. U-Net was first created for biomedical image segmentation, but it has now found extensive use in many other sectors where precise and thorough picture segmentation is necessary. U-Net's architecture is built to produce accurate segmentation even with a small number of training examples. Its unusual "U-shaped" construction is what gave rise to its moniker [37]. The contraction path (down sampling) and the expansion path (up sampling) comprise the two primary components of the U-Net Design-Net modifications and enhancements have been suggested to address particular issues or boost performance even more [38]. Research is still being conducted to improve U-Net's performance and capability for a variety of difficult segmentation tasks, as it continues to be a core architecture in image segmentation.

An improved variant of the original U-Net model, the Residual U-Net architecture uses residual connections to enhance segmentation performance [39]. These connections also referred to as skip connections address the disappearing gradient issue and make it possible to train deeper networks by improving the gradients' ability to move through the network. The U-Net architecture is strengthened and able to learn more complicated characteristics without appreciably raising the computational load thanks to the use of residual connections. The general U-shaped structure of the original U-Net model is maintained by residual U-Net, which consists of skip connections connecting the two paths as well as an expansion (decoder) path and contraction (encoder) path. The main innovation of Residual U-Net is that residual blocks are included to the expansion and contraction paths. Every residual block is made up of two or more convolutional layers with ReLU activations, and its input and output are combined via a shortcut connection. By reducing the degradation issue, this architecture helps the network learn residual functions in relation to the layer inputs, which makes it easier to train deeper networks. The skip connections that concatenate feature mappings from the contraction path to the appropriate layers in the expansion path are a distinguishing feature of residual U-Net. These connections are carried over from the original U-Net. For more precise segmentation, these skip connections are essential for fusing semantic data from deep layers with spatial data from shallow layers [40].

By fusing the advantages of residual learning with the advantages of the original U-Net model, the Residual U-Net architecture signifies a substantial breakthrough in the field of picture segmentation. Residual U-Net makes it possible to train deeper and more efficient segmentation models, which in turn enables more precise and thorough image analysis for a variety of applications, including medical imaging [41].

3. Proposed Method

Using a unique approach to convolutional neural network construction, the "Attention-Attention ResNet" model doubles the power of attention mechanisms inside U-Net framework. The goal of this integration as shown in Fig.1 is to greatly improve the model's capacity to concentrate on pertinent features for segmentation and classification tasks, especially in fields that need high precision, like medical imaging. By adding two attention processes to a residual network, the "Attention-Attention ResNet" (MAED U-Net) aims to address the drawbacks of conventional CNNs. Because of this design decision, the model may achieve previously unheard-of levels of concentration and feature refinement, which helps it perform better across a range of image analysis tasks.

3.1 MAED U-NET

The model is based on the U-Net architecture, which is well known for using residual blocks to train incredibly deep networks. These blocks offer shortcut connections that bypass one or more levels, which helps to mitigate the disappearing gradient issue. By focusing on the input's spatial dimensions, the first layer of attention helps the model better understand and identify patterns by allowing it to focus on specific regions of a picture that are important for the task at hand. After spatial refinement, each feature channel's relevance is assessed by the channel attention mechanism, which enables the model to dynamically refocus on the most informative features and enhances the model's capacity to discern subtleties in the data.

The attention gate is comparable to the human visual attention mechanism, which recognizes important feature information for a given task, highlights salient feature information in feature maps, and automatically focuses on the target region. It also learns to suppress irrelevant feature responses. Recent studies have demonstrated that networks function better when deep learning models trained with the attention gate implicitly.

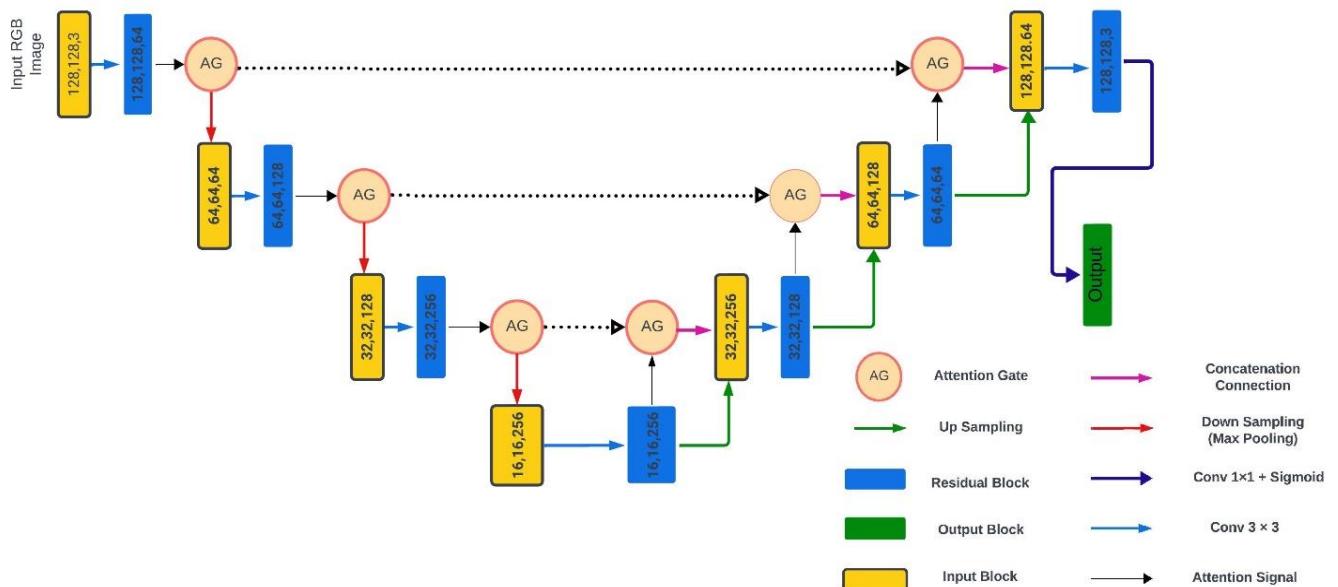
3.2 Attention Gate

A gating signal vector g_i is used to choose the focus regions from a coarser scale for each pixel i , where x_l is the feature map of the l layer. An illustration of an attention gate unit's typical architecture may be seen in Fig. 2 and Fig.

3. The

Figure 1 : MAED U-Net

attention coefficient, denoted by α , is a number between 0 and 1. It exports feature responses associated with the job at hand and suppresses irrelevant feature information. The element-wise multiplication of x_l and α results in the output



x_{out} , whose precise formula is shown in equation (1).

$$x_{out} = xl \cdot ai \quad (1)$$

Here, additive attention—as opposed to multiplicative attention—is used to determine the gating coefficient α . While adding attention comes at a considerable computational expense, the results of segmentation can be more promising. We use multi-dimensional attention coefficient to concentrate on a subset of target locations because skin lesion segmentation is a multiple semantic classes task. One way to compute the multi-dimensional attention coefficient is as follows:

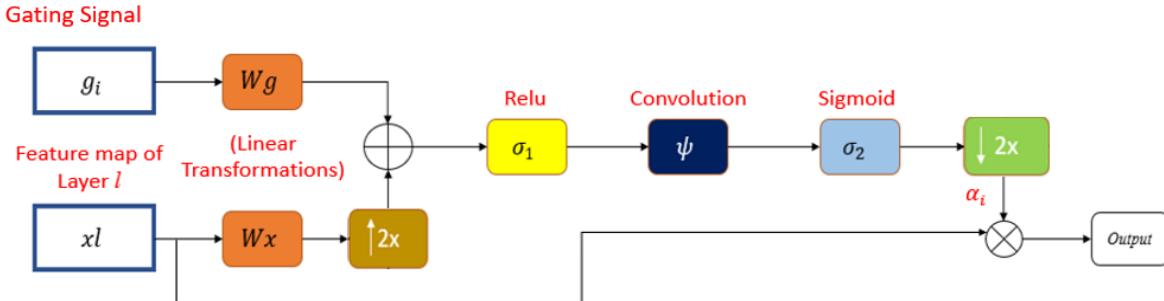
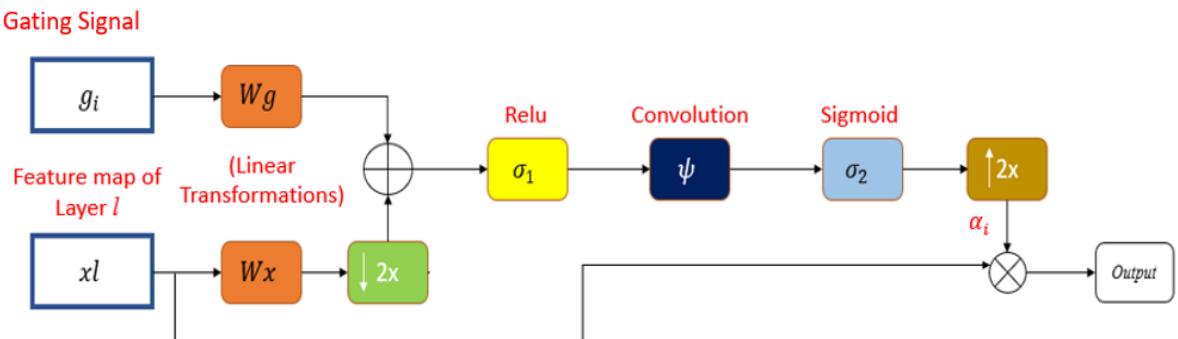


Figure 2: Attention Gate for Encoder

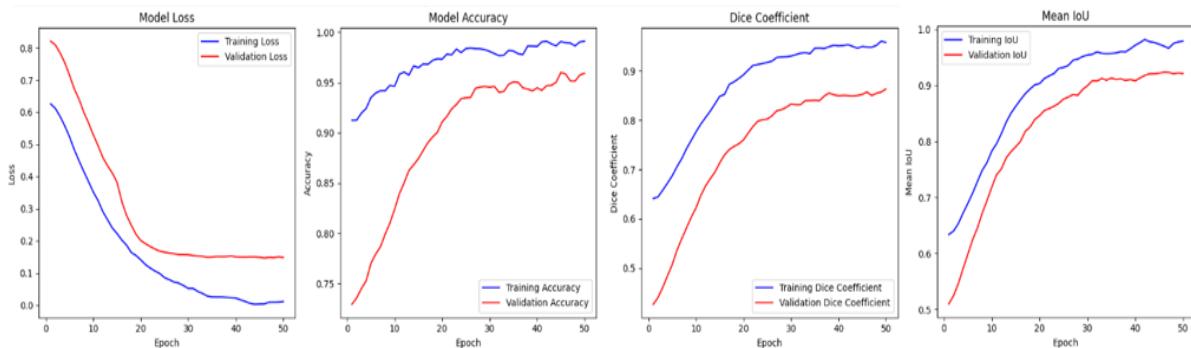


Here,
additive
attention—
as opposed to

multiplicative attention—is used to determine the gating coefficient α . While adding attention comes at a considerable computational expense, the results of segmentation can be more promising. We use multi-dimensional attention coefficient to concentrate on a subset of target locations because skin lesion segmentation is a multiple semantic classes task. One way to compute the multi-dimensional attention coefficient is as follows:

$$\alpha_i = \sigma_2 \left(\psi^T \left(\sigma_1 \left(W_x^T x_l + W_g^T g_i + b_\psi \right) \right) + b_\alpha \right) \quad (2)$$

where σ_1 is often chosen as ReLU function $\sigma_1(r) = \max(0; r)$, and σ_2 is the Sigmoid function described in equation (3). W_x , W_g and are linear transformations, and b_ψ and b_α are bias terms. We utilize 1×1 channel-wise convolution to perform linear transformation on the feature map x_l and the gating signal vector g_i .



$$\sigma_2(r) = \frac{1}{1+e^{-r}} \quad (3)$$

Higher accuracy in both classification and segmentation tasks can be achieved by AA: Res U-Net through the efficient implementation of attention mechanisms, which improves the network's ability to recognize and segment pertinent

features in complicated images. The model is appropriate for large-scale deployment and real-time applications because it retains computational efficiency, a feature of the UNet architecture, despite its complexity.

The AA: Res U-Net Model, which combines the benefits of residual learning and attention mechanisms with the tried-and-true U-Net architecture, is a major breakthrough in medical picture segmentation. This model solves some of the frequent issues with deep learning models, like training depth and feature emphasis, in addition to increasing the accuracy and efficiency of segmentation tasks. By utilizing this approach on datasets such as ISIC 2018 and HAM10000, researchers can expand the limits of automated medical diagnosis and treatment planning.

4. Results and Analysis

All the experiments including image pre-processing, tiling, training, validation and testing of the datasets were performed on open-source NVIDIA TESLA P100 GPU. All the programming was done in Python language using PyTorch.

We trained our model for 50 epochs and a batch size of 8. Image size for HAM10000 data set is 256. It can be inferred from the given below graphs in Fig 4 generated after the validation, that our model has achieved validation accuracy of 97.2%, validation loss of 15%, DSC 92.6% and mIoU 91%.

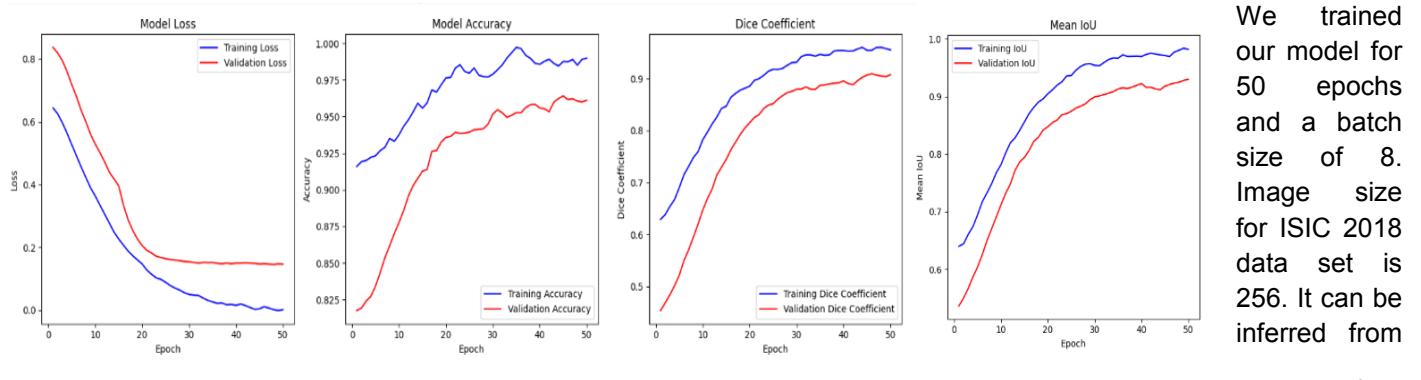


Figure 4: HAM10000 Performance Graphs

given below graphs in Fig 5 generated after the validation, that our model has achieved validation accuracy of 95.8%, validation loss of 18%, DSC 84.2% and mIoU 89%.

The comparison research in the table 1 shows that MAED U-Net performs better than the other model in terms of accuracy on both datasets, demonstrating the use of its residual learning and attention processes for the identification and segmentation of skin cancer lesions. The consistent performance of MAED U-Net, particularly in attaining the greatest accuracy, suggests that it has the potential to enhance diagnostic procedures by better segmenting dermatoscopic images. But the subtle variations in DSC, especially in the HAM10000 sample, point to regions that could have more improvement. Even while MAED U-Net performs exceptionally well in terms of overall accuracy, clinical applications depend on the accuracy of segmentation borders, as demonstrated by DSC. These findings not only support MAED U-Net's novel methodology, but they also open the door for further developments and uses in the field of medical image analysis.

Figure 5: ISIC 2018 Performance Graphs

Table 1: Comparative Analysis of our Model with SOTA Technologies

Architecture	Year	Data Set	Accuracy
U-Net [45]	2019	HAM10000	0.924
Res- Net [45]	2019	HAM10000	0.935

MAG-UNet [46]	2024	HAM10000	0.963
SLED [50]	2018	ISIC 2018	0.869
U-Net [45]	2019	ISIC 2018	0.887
Res-Net [45]	2019	ISIC 2018	0.893
ADAM [49]	2020	ISIC 2018	0.947
MobileNetV3-UNet [48]	2021	ISIC 2018	0.947
Attn_U_Net-GN [47]	2023	ISIC 2018	0.951

5. Segmented Output Results

Fig.6 depicts the segmented output for HAM10000 data set and Fig.7 depicts the segmented output for ISIC 2018 Dataset of our proposed model.

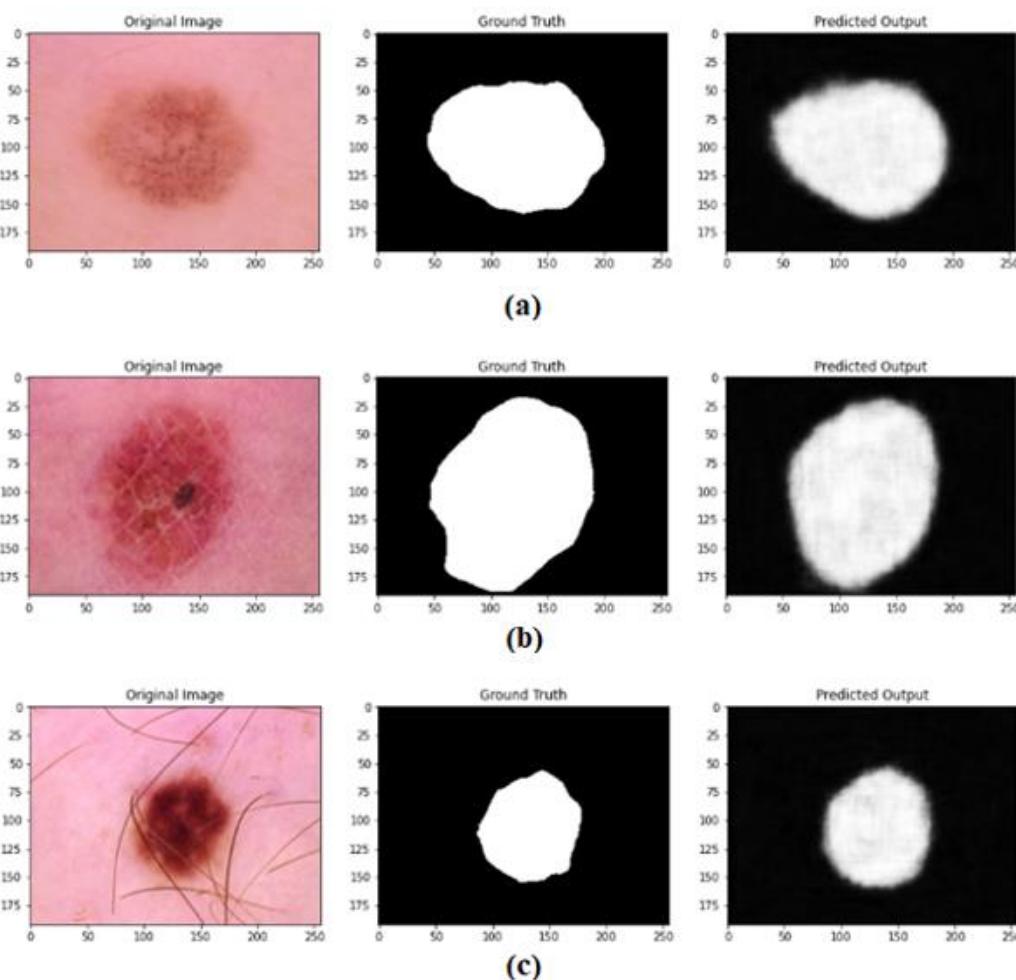


Figure 6: Segmentation Results for HAM10000 (a) Melanoma (b) Basal Cell Carcinoma (c) Actinic Keratosis

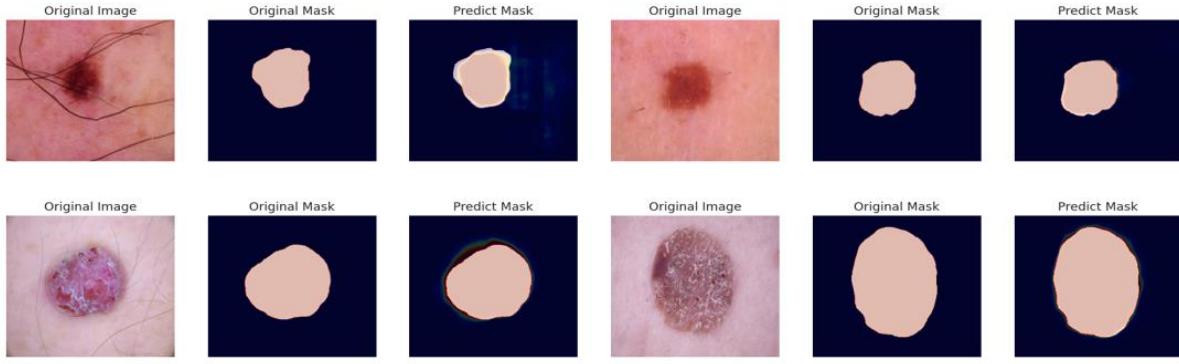


Figure 7: Segmentation Results for ISIC 2018

The summary of the results achieved by our proposed model is depicted in Table 2

Data Set	Loss	Accuracy	Dice Score	μ IoU
HAM10000	0.15	0.972	0.926	0.91
ISIC 2018	0.18	0.958	0.842	0.89

6. Conclusions and Future Direction

This study effectively created and assessed a novel deep learning model called MAED U-Net for the segmentation of skin cancer lesions from dermoscopic pictures. Our comparative study between U-Net and Residual U-Net on two important datasets, HAM10000 and ISIC 2018, shows how well MAED U-Net performs. MAED U-Net has demonstrated its effectiveness in correctly segmenting skin cancer lesions, with accuracy and Dice Similarity Coefficient (DSC) scores exceeding those of the benchmark models. Our proposed model lacks behind in the accuracy achieved by the latest state of the art technologies for HAM10000 data set as discussed in section 4.6 because we have used unbalanced HAM10000. In future if we use balanced HAM1000, there is a likely chance that we can outperform the MAG-ResNet. This highlights the tool's potential as a significant tool in the early identification and diagnosis of skin cancer.

MAED U-Net excels because of its creative incorporation of attention processes into a residual network structure, which allows for accurate focus on pertinent features and efficient resolution of the vanishing gradient problem—a frequent difficulty in deep network training. These technological developments add to the model's exceptional accuracy and ability to greatly improve dermatology diagnostic processes.

Notwithstanding its successes, the study notes its shortcomings, including the model's computing requirements and reliance on the caliber and variety of training data. These variables provide crucial information about areas that need more investigation and improvement.

There are numerous and potential avenues for future investigation. Priority will be given to improving the model's generalizability by adding more broad and diverse datasets, which may include underrepresented skin varieties and lesser-known forms of skin cancer. Moreover, enhancing MAED U-Net for real-time clinical application is an important objective with the goal of increasing accessibility to effective and precise skin cancer diagnostics.

A route to democratizing access to early skin cancer detection techniques may involve investigating the integration of MAED U-Net with mobile health technology, particularly in areas with a dearth of dermatological care. Furthermore, expanding the applicability of MAED U-Net's fundamental ideas to additional medical imaging segmentation jobs could increase its influence and even revolutionize other healthcare domains' diagnostic procedures.

To sum up, MAED U-Net is a potent tool for the early identification and diagnosis of skin cancer and represents a significant achievement in the field of medical picture segmentation. Its creation not only highlights how deep learning can be used to improve medical diagnoses, but it also paves the way for future advancements that should lead to better patient outcomes and treatment.

7. References

- [1] S.-T. Tran, C.-H. Cheng, T.-T. Nguyen, M.-H. Le, and D.-G. Liu, "TMD-UNet: Triple-UNet with multi-scale input features and dense skip connection for medical image segmentation," *Healthcare*, vol. 9, no. 1, p. 54, Jan. 2021.
- [2] Nh Narayana Health, Health for All, All for Health, NH CARES Brain Tumour, Types, Risk Factors, Symptoms, and Surgery. Accessed: Nov. 10, 2023. [Online]. Available: <https://www.narayanahealth.org/diseases/braintumour>
- [3] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.
- [4] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, no. 13, pp. 97–129, Jul. 2013.
- [5] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
- [6] S. Iqbal, M. U. Ghani, T. Saba, and A. Rehman, "Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN)," *Microsc. Res. Technique*, vol. 81, no. 4, pp. 419–427, Apr. 2018.
- [7] Robin, Mirya, Jisha John, and Aswathy Ravikumar. "Breast tumor segmentation using U-NET." 2021 5th international conference on computing methodologies and communication (ICCMC). IEEE, 2021.
- [8] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. Technol.*, vol. 19, no. 1, pp. 252–258, Feb. 2013.
- [9] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: UNet-like pure transformer for medical image segmentation," in Proc. Eur. Conf. Compute. Vis., Tel Aviv, Israel. Berlin, Germany: Springer, Oct. 2022, pp. 205–218.
- [10] Z. Rauf, A. Sohail, S. H. Khan, A. Khan, J. Gwak, and M. Maqbool, "Attention-guided multi-scale deep object detection framework for lymphocyte analysis in IHC histological images," *Microscopy*, vol. 72, no. 1, pp. 27–42, Feb. 2023.
- [11] M. L. Ali, Z. Rauf, A. R. Khan, and A. Khan, "Channel boosting based detection and segmentation for cancer analysis in histopathological images," in Proc. 19th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Aug. 2022, pp. 1–6.
- [12] S. H. Khan, A. Khan, Y. S. Lee, M. Hassan, and W. K. Jeong, "Segmentation of shoulder muscle MRI using a new region and edge based deep auto-encoder," *Multimedia Tools Appl.*, vol. 82, no. 10, pp. 14963–14984, Apr. 2023.
- [13] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020.
- [14] Robin, Mirya, Jisha John, and Aswathy Ravikumar. "Breast tumor segmentation using U-NET." 2021 5th international conference on computing methodologies and communication (ICCMC). IEEE, 2021.
- [15] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.
- [16] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
- [17] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, arXiv:1804.03999.
- [18] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using markercontrolled watershed, tracking using mean-shift, and Kalman filter in timelapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2405–2414, Nov. 2006.
- [19] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS One*, vol. 8, no. 7, Jul. 2013, Art. no. e70221.
- [20] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 137–154, Feb. 2003.
- [21] J. Walsh, A. Othmani, M. Jain, and S. Dev, "Using U-Net network for efficient brain tumor segmentation in MRI images," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100098.
- [22] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-UNet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

- [23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in Proc. Deep Learn. Medical Image Anal. Multimodal Learn. Clin. Decision Support, 4th Int. Workshop, 8th Int. Workshop, Granada, Spain. Berlin, Germany: Springer, 2018, pp. 3–11.
- [24] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," Skin Res. Technol., vol. 19, no. 1, pp. 252–258, Feb. 2013.
- [25] F. Peruch, F. Bogo, M. Bonazza, V.-M. Cappelleri, and E. Peserico, "Simpler, faster, more accurate melanocytic lesion segmentation through MEDS," IEEE Trans. Biomed. Eng., vol. 61, no. 2, pp. 557–565, Feb. 2014.
- [26] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," Comput. Methods Programs Biomed., vol. 162, pp. 221–231, Aug. 2018.
- [27] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional deconvolutional networks," IEEE J. Biomed. Health Informat., vol. 23, no. 2, pp. 519–526, Mar. 2019.
- [28] Ş. Öztürk and U. Özkaray, "Skin lesion segmentation with improved convolutional neural network," J. Digit. Imag., vol. 33, no. 4, pp. 958–970, Aug. 2020.
- [29] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network," Computing Methods Programs Biomed., vol. 186, Apr. 2020, Art. no. 105241.
- [30] X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, and P. Wu, "ASCU-Net: Attention gate, spatial and channel attention U-Net for skin lesion segmentation," Diagnostics, vol. 11, no. 3, p. 501, Mar. 2021.
- [31] S. Kadry, D. Taniar, R. Damaševicius, V. Rajinikanth, and I. A. Lawal, "Extraction of abnormal skin lesion from dermoscopy image using VGG-SegNet," in Proc. 7th Int. Conf. Bio Signals, Images, Instrum. (ICBSII), Mar. 2021, pp. 1–5.
- [32] V. Rajinikanth, S. Kadry, R. Damaševicius, D. Sankaran, M. A. Mohammed, and S. Chander, "Skin melanoma segmentation using VGG-UNet with Adam/SGD optimizer: A study," in Proc. 3rd Int. Conf. Intell. Comput. Instrum. Control Technol. (ICICICT), Aug. 2022, pp. 982–986.
- [33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2818–2826.
- [34] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," IEEE Trans. Med. Imag., vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
- [35] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [36] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer," in Proc. AAAI Conf. Artif. Intell., Jun. 2022, vol. 36, no. 3, pp. 2441–2449.
- [37] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.
- [38] W. Chi, L. Ma, J. Wu, M. Chen, W. Lu, and X. Gu, "Deep learning-based medical image segmentation with limited labels," Phys. Med. Biol., vol. 65, no. 23, Dec. 2020, Art. no. 235001.
- [39] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," IEEE Trans. Med. Imag., vol. 39, no. 10, pp. 3008–3018, Oct. 2020.
- [40] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, arXiv:1706.05587.
- [41] Gómez-de-Mariscal, Estibaliz, et al. "Deep-learning-based segmentation of small extracellular vesicles in transmission electron microscopy images." Scientific reports 9.1 (2019): 13211.
- [42] D. Misra, "Mish: A self-regularized non-monotonic activation function," 2019, arXiv:1908.08681.
- [43] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, arXiv:1711.05101.
- [44] S. Dev, H. Javidnia, M. Hossari, M. Nicholson, K. McCabe, A. Nautiyal, C. Conran, J. Tang, W. Xu, and F. Pitié, "Identifying candidate spaces for advert implantation," in Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT), Oct. 2019, pp. 503–507.
- [45] Kaur, Rajdeep, and Sukhjeet Kaur. "Automatic skin lesion segmentation using attention residual U-Net with improved encoder-decoder architecture." Multimedia Tools and Applications (2024): 1-27.
- [46] Hussain, Tahir, and Hayaru Shouno. "MAGRes-UNet: Improved Medical Image Segmentation through a Deep Learning Paradigm of Multi-Attention Gated Residual U-Net." IEEE Access (2024).

- [47] Arora R, Raman B, Nayyar K, Awasthi R (2021) Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomed Signal Process Control* 65:102358. <https://doi.org/10.1016/j.bspc.2020.102358>
- [48] Wibowo A, Purnama SR, Wirawan PW, Rasyidi H (2021) Lightweight encoder-decoder model for automatic skin lesion segmentation. *Inf Med Unlocked* 25:100640. <https://doi.org/10.1016/j imu.2021.100640>
- [49] Wu H, Pan J, Li Z, Wen Z, Qin J (2020) Automated skin lesion segmentation via an adaptive dual attention module. *IEEE Trans Med Imaging* 40(1):357–370. <https://doi.org/10.1109/TMI.2020.3027341>
- [50] Zeng G, Peng H, Li A, Liu Z, Liu C, Yu PS, He L (2023) Unsupervised skin lesion segmentation via structural entropy minimization on multi-scale superpixel graphs. *arXiv preprint arXiv:2309. 01899*. <https://doi.org/10.48550/arXiv.2309.01899>

Article Citation:
Khalil et al. (2024). “
Frequent Subgraph Mining: Opportunities, Challenges and Future Prospects”.
FUSST International Conference on Computer Science and Engineering

Frequent Subgraph Mining: Opportunities, Challenges and Future Prospects

Muhammad Ibrahim Khalil ^{a,*}, Saif Ur Rehman ^a

^a University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan
m.ibrahim.khalil@outlook.com, saif@uaar.edu.pk

* Corresponding author: m.ibrahim.khalil@outlook.com

Abstract:

Graphs have grown more significant in modeling and visualizing data. Graph considered as a key area of study in the field of data mining is graph mining. The research focuses on identifying common subgraphs in graph data collections. Graphs are widely used data structures for modeling real-world processes. Graph Mining is a subset of data mining in which large amounts of complicated data are presented as graphs and mined to extract information. The method of mining data sets depicted as graph structures, known as graph mining, is extensively researched in biological informatics, computer systems, social media, etc. Frequent subgraph mining is an area of the graph mining space that is widely utilized for graph classification, index construction, and graph clustering. The frequent subgraph mining is treated from numerous viewpoints and regarded in different ways depending on the domain predictions. This article surveys the latest studies on the topic of frequent subgraph mining and demonstrates solutions to the major research concerns.

Keywords: Graph Mining, Frequent subgraph mining, Social Media, Frequent Subgraph Patterns



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The graph is a data structure that represents relationships or connections between entities or data points. Graphs in data mining can take various forms, and they are used to model and analyze complex relationships and patterns within datasets [1, 2]. In recent years, there has been a significant amount of research conducted on graph mining with the aim of improving performance and developing advancement [3]. An innovative approach involves the extraction of data from graphs, which presents a novel challenge [4]. The representation of a structure is determined by its proper interactions, and a graph is a suitable way to depict these relations [5]. Consequently, the process of acquiring insights from graph-structured data shows a broad challenge for extracting useful information [6].

Graph mining (GM) is widely used in several domains, including the exploration of molecular substructures, analysis of online links, identification of outliers, examination of chemical compounds, and investigation of social networks [7, 8]. The exploration of frequent subgraph mining has significant importance within the field of graph mining [9, 10]. The objective of the mining process is to identify and extract all frequent subgraphs within a given collection of graphs [11]. Several methods have been suggested for the purpose of mining frequent subgraphs. However, many of these techniques are not suitable for handling huge datasets due to scalability limitations [12, 13]. The network diagram for social networks is shown in Figure 1.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

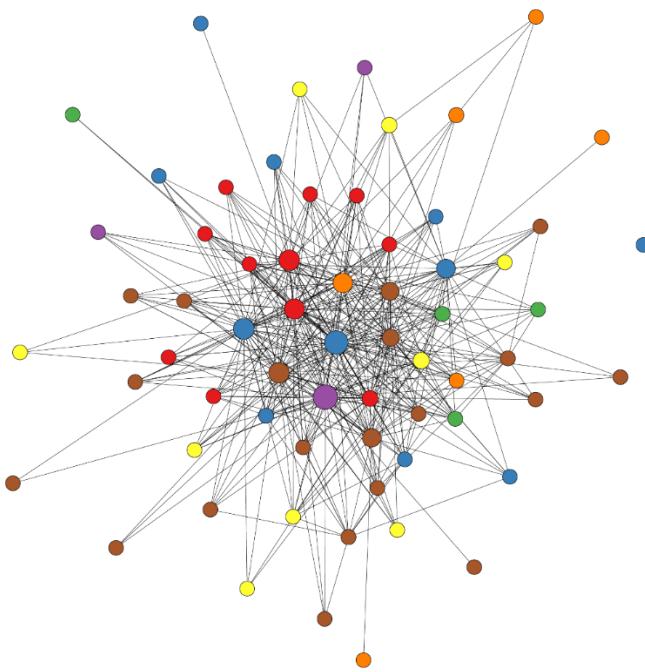


Figure 1: Network graph for social networks

The identification of recurring patterns within a database, specifically patterns that manifest in a minimum defined number of database elements, is a fundamental objective explored within the field of data mining [14]. Frequent patterns possess inherent interest and can serve as valuable foundations or features for predicting analyses of data [15, 16]. Due to the importance of application domains such as molecules of chemicals analysis or graphical representations on the global web, there has been a growing interest in algorithms capable of efficiently discovering frequent patterns in databases containing structured objects [17].

By systematically identifying and analyzing recurring substructures within complex graphs or datasets, frequent subgraph mining enables the extraction of essential patterns and associations [18]. This process supports in unveiling of critical insights across diverse applications, ranging from chemical compound analysis to social network understanding and web page recommendation [19, 20]. In essence, frequent subgraph mining empowers data mining practitioners to unravel the underlying structure and significance within their data, leading to more informed decision-making, enhanced predictive models, and a deeper understanding of complex relationships [21, 22]. It plays an essential role in uncovering valuable knowledge that drives innovation and informed decision-making in a wide range of domains [23, 24]. FSM presents a number of challenges that make it difficult to identify recurring structures in big graph datasets. As the size of these databases grows, frequent subgraph identification becomes more difficult. Thus, the field requires novel approaches that not only improve the efficiency of frequent subgraph mining, but also identify significant subgraph patterns.

In this paper, we presented a complete overview of the various graph mining methodologies. A detailed description of each of these approaches has been provided, including information on their respective techniques, significant research contributions, and limitations. These approaches have been extensively reviewed. The remaining content of this paper is arranged as follows: Section II discusses the primary terminology employed in graph theory. Section III provides a comprehensive literature overview of graph mining approaches developed in the previous several decades. Section IV addresses the critical inquiry of these various graph mining algorithms, the specifics of which are provided in section II. This research accomplishes the findings of our work and potential future paths in Section V.

2. Literature Review

Existing subgraph mining techniques are not effective when dealing with huge networks, such as those seen in social areas. Furthermore, these graphs are often generated as a consequence of an ongoing procedure, thus they are subjected to ongoing change. In social media platforms, user interactions contribute new edges to the graph construction, which is constantly evolving. When the graph modifies by adding or deleting an edge, a substantial amount

of new subgraphs are produced, and existing subgraphs might be updated or deleted. Keeping track of all potential graph modifications leads to high computational cost.

Table 1: Recent Graph Mining Approaches

Article	Technique	Dataset	Research Contribution
[25]	TANN (Triple Attentive Neural Network)	DIGINETICA	Addresses the problem of session-based recommendations.
[26]	Mining code change patterns	Open-source code repositories	Captures context relations between individual edits and extracts frequent subgraphs
[27]	GNCCP (Generalized Network Connectivity Constraint Problem)	Vehicular networks dataset	Indicates the effectiveness in extracting stable subgraphs.
[28]	FS3 Change	Java projects on GitHub	Method for mining change patterns in software repositories.
[29]	FSM with graph sampling	MICO	A graph sampling function to improve the efficiency of the GraMi.
[30]	SampleMine	SiteSeer, MiCo, Orkut, Friendster	Introduced a nested loop and loop perforation, resulting in considerable speedups.

From the literature review, it was found that FSM in large-scale graphs requires algorithms capable of efficiently handling massive amounts of data. Identifying meaningful subgraph patterns amidst noise and irrelevant structures in the graph is challenging. Storing and processing large graphs in memory can be challenging due to limited memory resources. Ensuring the quality and relevance of mined frequent subgraphs is essential for practical applications. There is a need to identify the significant patterns from the discovered subgraphs. A significant pattern is often defined as a subgraph whose discovery provides deep and important insights into the underlying data, revealing fundamental structural or relational qualities that have a significant business impact.

3. Mining of Frequent Subgraph Patterns

Subgraph mining involves searching for recurring patterns of connected nodes within a network representing relationships between entities such as users or subjects. The process usually involves finding clusters of nodes or subgraphs that share characteristics or play important roles in the network. Subgraph mining algorithms are used to identify relevant patterns and relationships within a network. One common technique is to use level-wise iterative algorithms, to search for similar size- k frequent subgraphs and merge them to form larger subgraphs of size $(k+1)$. These algorithms efficiently identify recurring network structures, allowing valuable insights to be extracted from social media data. The figure. 2 and algorithm 1 depicts the subgraph mining workflow.

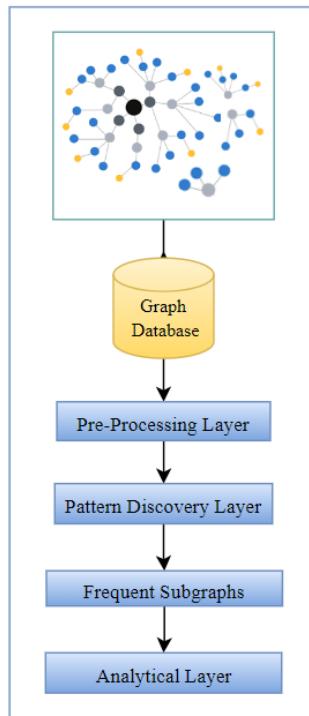


Figure 2: Graph mining technique working flow

In the context of social media, these subgraphs may represent recurring patterns of user connections, popular topics of discussion, or influential clusters within the network.

We will propose a significant subgraph mining approach for the different modules to extract the substantial patterns. Analyzing these subgraphs can provide researchers and analysts with valuable insights into user behavior, popular trends, and influential entities medical domain or social media platforms.

Additionally, metrics and principles are used to assess the significance and interest of the discovered subgraphs. Subgraph mining in social media is critical for identifying meaningful patterns, connections, and communities within the network, resulting in valuable insights for a variety of applications such as targeted advertising, content recommendation, and trend analysis.

Algorithm 1: Extract Frequent Subgraph Patterns

```

function mine_frequent_subgraphs(graph, min_sup):
1: max_size = determine_max_size(graph)
2: for size in range(1, max_size + 1):
    2.1. subgraphs = enumerate_subgraphs(graph, size)
    2.1.1 subgraphs = [list]
    2.1.2 for node in graph.nodes:
        2.1.3 subgraph = dfs_subgraph(graph, node, size)
        2.1.4 if subgraph:
            2.1.5 subgraphs.append(subgraph)
    2.1.5 return subgraphs
    2.2. for subgraph in subgraphs:
        2.2.1. support = count_support(graph, subgraph)
        2.2.2. if support >= min_sup:
            2.2.4. frequent_subgraphs[subgraph] = support
    3.4. return frequent_subgraphs

```

4. Conclusion and Future Work

Graph based mining algorithms concentrate on the extraction of frequent subgraphs in all domains. Subgraph extraction approaches are concerned with the detection of patterns in graphs that reflect any particular network structure which is considered useful among the set of provided data. In our study, we investigated frequent subgraph mining approaches, aiming to provide a comprehensive overview of current research from diverse perspectives while incorporating the most pertinent algorithms in the field. FSM is geared towards uncovering all subgraphs within a larger graph where occurrences meet a specified frequency threshold. Given the occurrence of extensive underlying graphs, particularly in domains such as medical, social media platforms in which efficiency is vital. Minimizing graph equivalence relation becomes crucial as it directly impacts computational efficiency, thereby enhancing the effectiveness of the mining operation. However, existing FSM algorithms primarily work for small datasets, there is a need for innovative concepts and approaches that enable to effectively handle large-scale datasets.

5. References

- [1] Bhavsar, S.A., V.H. Patil, and A.H. Patil, *Graph partitioning and visualization in graph mining: a survey*. Multimedia Tools and Applications, 2022. 81(30): p. 43315-43356.
- [2] Kang, J., et al. *Inform: Individual fairness on graph mining*. in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2020.
- [3] Jamshidi, K., R. Mahadasa, and K. Vora. *Peregrine: a pattern-aware graph mining system*. in *Proceedings of the Fifteenth European Conference on Computer Systems*. 2020.
- [4] Kang, J. and H. Tong. *Fair graph mining*. in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021.
- [5] Chen, Q., B. Tian, and M. Gao. *FINGERS: exploiting fine-grained parallelism in graph mining accelerators*. in *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*. 2022.
- [6] Hassan, A.K. and S.N. Mohammed, *A novel facial emotion recognition scheme based on graph mining*. Defence Technology, 2020. 16(5): p. 1062-1072.
- [7] Akhtar, N. and M.V. Ahamed, *Graph tools for social network analysis*, in *Research Anthology on Digital Transformation, Organizational Change, and the Impact of Remote Work*. 2021, IGI Global. p. 485-500.
- [8] Marmo, R., *Social media mining*, in *Encyclopedia of organizational knowledge, administration, and technology*. 2021, IGI Global. p. 2153-2165.
- [9] Liao, X., D. Zheng, and X. Cao, *Coronavirus pandemic analysis through tripartite graph clustering in online social networks*. Big Data Mining and Analytics, 2021. 4(4): p. 242-251.
- [10] Fournier-Viger, P., et al., *A survey of pattern mining in dynamic graphs*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2020. 10(6): p. e1372.
- [11] Bian, T., et al. *Rumor detection on social media with bi-directional graph convolutional networks*. in *Proceedings of the AAAI conference on artificial intelligence*. 2020.
- [12] Peng, H., et al., *Introduction to the Special Issue on Advanced Graph Mining on the Web: Theory, Algorithms, and Applications: Part 1*. 2023, ACM New York, NY. p. 1-2.
- [13] Lettieri, N., et al., *Knowledge mining and social dangerousness assessment in criminal justice: metaheuristic integration of machine learning and graph-based inference*. Artificial Intelligence and Law, 2023. 31(4): p. 653-702.
- [14] Li, J., et al., *Adaptive Subgraph Neural Network with Reinforced Critical Structure Mining*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
- [15] Reddy, A.S., et al., *Mining subgraph coverage patterns from graph transactions*. International Journal of Data Science and Analytics, 2022: p. 1-17.
- [16] Preti, G., G. De Francisci Morales, and M. Riondato, *Maniacs: Approximate mining of frequent subgraph patterns through sampling*. ACM Transactions on Intelligent Systems and Technology, 2023. 14(3): p. 1-29.

- [17] Nguyen, L.B., et al., *An efficient and scalable approach for mining subgraphs in a single large graph*. Applied Intelligence, 2022. 52(15): p. 17881-17895.
- [18] Jamshidi, K., H. Xu, and K. Vora. *Accelerating Graph Mining Systems with Subgraph Morphing*. in *Proceedings of the Eighteenth European Conference on Computer Systems*. 2023.
- [19] Yuan, L., et al., *T-FSM: A Task-Based System for Massively Parallel Frequent Subgraph Pattern Mining from a Big Graph*. Proceedings of the ACM on Management of Data, 2023. 1(1): p. 1-26.
- [20] Lim, S., et al., *Supervised chemical graph mining improves drug-induced liver injury prediction*. iScience, 2023. 26(1).
- [21] Nguyen, T.T., et al., *Scalable Maximal Subgraph Mining with Backbone-Preserving Graph Convolutions*. Information Sciences, 2023: p. 119287.
- [22] Yang, K., et al. *Extract and refine: Finding a support subgraph set for graph representation*. in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023.
- [23] Wu, Y., et al. *Shogun: A Task Scheduling Framework for Graph Mining Accelerators*. in *Proceedings of the 50th Annual International Symposium on Computer Architecture*. 2023.
- [24] Kanti Kumar, P., A. Dutta, and P. Kumar, *Application of Graph Mining Algorithms for the Analysis of Web Data*. Anurag and Kumar, Padmanavan, Application of Graph Mining Algorithms for the Analysis of Web Data (February 21, 2023), 2023.
- [25] Khan, F., et al., *Context-aware and Click Session-based Graph Pattern Mining with Recommendations for Smart EMS through AI*. IEEE Access, 2023.
- [26] Janke, M. and P. Mäder, *Graph based mining of code change patterns from version control commits*. IEEE Transactions on Software Engineering, 2020. 48(3): p. 848-863.
- [27] Meng, Y., et al., *Graph Similarity-Based Maximum Stable Subgraph Extraction of Information Topology From a Vehicular Network*. IEEE Transactions on Intelligent Transportation Systems, 2020. 23(1): p. 355-367.
- [28] Janke, M. and P. Mäder, *FS 3 change: A Scalable Method for Change Pattern Mining*. IEEE Transactions on Software Engineering, 2023.
- [29] Sahu, S., et al., *Mining approximate frequent subgraph with sampling techniques*. Materials Today: Proceedings, 2023. 81: p. 395-402.
- [30] Jiang, P., et al. *SampleMine: A Framework for Applying Random Sampling to Subgraph Pattern Mining through Loop Perforation*. in *Proceedings of the International Conference on Parallel Architectures and Compilation Techniques*. 2022.

Article Citation:
Mudasar et al. (2024).
Canned Apple Fruit
Freshness Detection Using
Hybrid Convolutional Neural
Network and Transfer
Learning". FUSST
International Conference on
Computer Science and
Engineering



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024
AuthorI_LastName et al.



Published by
Foundation University
Islamabad.

Web: <https://fui.edu.pk/>

Canned Apple Fruit Freshness Detection Using Hybrid Convolutional Neural Network and Transfer Learning

Mudasar Iqbal^a Syed Tehseen Haider^{a,*} Saif Ur Rehman^b

^a School of Software Engineering, Xinjiang University, Urumqi, China

^b University Institute of Information Technology, PMAS-Arid Agriculture University Rawalpindi, Rawalpindi, Pakistan.

* Corresponding author: Syedtahseenh@gmail.com

Abstract:

Fruits contain good nutrients such as protein and vitamins, therefore fruits are usually used as complementary food ingredients. Apple fruit is one of the most important traditional table fruits in the temperate zone besides being the most commonly consumed fruit in the world. Apple's freshness is a decisive characteristic of consumer choice. Consumption may increase if apples of satisfactory freshness and quality are provided to the consumers. Apple fruits with less freshness cause major agricultural problems and economic loss. Many approaches have been presented to predict apple fruit freshness, however, most of these studies suffer from low accuracy due to poor image quality and the unavailability of sufficient data. This study aims to grade the freshness level of apple fruits by combining digital image processing and machine learning techniques. The proposed technique involves data collection of apple fruit, image segmentation, and preprocessing, creating the neural network model, training the network model, and testing for apple freshness prediction. A dataset is collected for apple fruit images for fresh, semi-fresh, and rotten classes which is further augmented for more images. We extensively evaluated the performance of the proposed convolutional neural network- transfer learning (CNN-TL)-based technique. For performance evaluation, ResNet, GoogleNet, AlexNet, VGG, MobileNetV2, and InceptionV3 are also utilized due to their reported superior performance. The experimental results show that the proposed approach achieved 98% accuracy on the freshness classification of the apple fruits, both on the original as well as on the augmented dataset.

Keywords: Image Processing; Convolutional Neural Network-Transfer Learning; Neural Network Model.

1. Introduction

Preserving fruit freshness is a critical issue for both consumers and the agricultural industry, as apple fruits are among the most widely consumed and economically important crops worldwide. The freshness of apples is crucial not only for maintaining consumer satisfaction, but also for reducing economic losses caused by spoilage and loss of quality. Reportedly, approximately one-third of harvested fruits, including apples, are wasted due to decay during storage and distribution [1] [2] [3]. Freshness is a crucial factor in consumers' purchasing decisions and plays an important role in market sales. When apples lose their freshness, they become unattractive and less nutritious, reducing their commercial value [4].

Traditionally, apple sorting is done manually, with trained professionals visually inspecting the fruit for signs of spoilage, damage or decay. However, this approach is labor-intensive, time-consuming, and often inconsistent due to human subjectivity [5]. Therefore, researchers and engineers are increasingly turning to automated solutions such as computer vision and machine learning algorithms to improve the accuracy and efficiency of apple grading systems. These technologies enable a more objective, reliable and scalable solution to the challenge of maintaining the freshness of apples throughout the supply chain [6]. The introduction of digital image processing (DIP) techniques and convolutional neural networks (CNNs) has shown promise for automating freshness and quality detection in fruits and vegetables [7]. By analyzing features such as color, texture and shape, these systems can accurately classify products based on their visual characteristics. However, many existing fruit freshness detection approaches still have limitations, including low accuracy due to insufficient image quality, insufficient training data, and difficulty in distinguishing between slightly degraded and completely spoiled fruits [8]. Furthermore, many existing machine learning models require large and diverse datasets to achieve high accuracy, which may not always be available for certain agricultural applications [9].

To address these issues, this study proposes a hybrid model that combines convolutional neural networks (CNNs) with transfer learning (TL) to improve the accuracy and efficiency of apple freshness detection. Transfer learning allows the

model to leverage pre-trained networks that have already learned useful features from large datasets and apply them to the specific task of apple classification [10] [11] [12]. This approach mitigates the challenge of limited data by leveraging knowledge from other domains, improving the model's generalization ability and reducing training time. The CNN-TL hybrid model developed in this study is specifically designed to classify apple fruits into three categories: fresh, semi-fresh and rotten. This classification is based on a dataset of apple images that have been segmented, preprocessed and augmented to ensure diversity and robustness [13].

The proposed system addresses the shortcomings of previous methods by using a combination of digital image processing and machine learning. First, the system collects apple fruit images from different freshness stages, including fresh, semi-fresh, and rotten, and augments these images to create a robust dataset for training. Next, image segmentation techniques such as K-means clustering are applied to the dataset to extract meaningful features, which are then used to train a CNN-TL model [14]. By comparing the performance of this model with other well-known architectures such as ResNet, GoogleNet, AlexNet, VGG, MobileNetV2 and InceptionV3, the study shows the superiority of the CNN-TL approach in terms of accuracy and efficiency [15] [16] [17].

The results of this research show that the proposed hybrid CNN-TL model achieves an impressive 98% accuracy in classifying apples based on their freshness. This is a significant improvement over previous approaches, which often struggled with accuracy due to poor image quality or insufficient data [18]. By using transfer learning, the model is able to make better predictions with a smaller dataset, making it a viable solution for real-world applications where large, diverse datasets may not be available. Furthermore, the model's high accuracy in detecting even minor differences between fresh and semi-fresh apples offers practical benefits to the agricultural industry as it can help reduce waste and ensure consumers receive high-quality products [19].

This study presents a novel solution to the apple freshness detection problem by leveraging the power of CNN and Transfer learning. The proposed model not only achieves high accuracy but also shows the potential for scalability and real-world application. This advancement represents a significant advance in the automation of fruit grading systems, offering providing a practical, efficient and accurate method of ensuring apple freshness throughout the supply chain. The implications of this research go beyond apples and could be applied to other fruits and agricultural products, paving the way for further innovation in food quality assessment.

2. Related Work

The study [20] presented an approach to determine the type of apple using a VGGNet network. First, a dataset of fruit images is used to train the VGGNet architecture. The YOLOv3 model is then loaded with weights that have already been trained. The apple freshness assessment model and the apple type detection model are the two separate models that make up the system. Preprocessing is done and then the VGGNet model is used to predict the type of apple. The image is then run through the YOLOv3 model to ascertain its freshness.

Researchers used real-world images with complex disturbances to investigate the detection and categorization of apple quality in [21]. The work presented a novel CNN-based model intended to accomplish accurate and quick apple quality ratings. The suggested model showed remarkable performance through rigorous training and validation, reaching peak training and validation accuracies of 99% and 98.98% at the 2590th and 3000th epochs, respectively. The suggested model's overall accuracy was 95.33% when evaluated on a separate dataset with 300 apples.

Using photographs of both fresh and rotting oranges, apples, and bananas in the dataset, the authors presented a method for classifying fruit freshness in [22]. The model performance was assessed using a variety of techniques, such as CNN, Inception, Xception, k nearest neighbor (KNN), support vector machine (SVM), random forest (RF), decision tree (DT), and Naive Bayes (NB). With an accuracy of 97.25%, the suggested method with the Xception algorithm yielded the highest accuracy, ahead of the Inception algorithm at 95.55% accuracy. CNN scored 77.17%, RF 77.17%, SVM 75.87%, KNN 61.3%, NB 59.38%, and DT showed a 44.29% accuracy. The study [23] examined three apple diseases: apple scab, apple rot, and apple blotch. It uses K-mean clustering for image segmentation and extracts features like global color histogram, color coherence vector, local binary pattern, and complete local binary pattern. The authors reported 93% classification accuracy.

Smart bundling based on the biochemical sensor is used to assess food quality and safety from creation to client. These sensors check different dietary parameters, for example, carbon dioxide, pH value, temperature, and temperature to monitor food quality and health. This exploration is significantly more centered on the well-being of food instead of checking the nature of food. It portrays various strategies to further develop our food pressing style called brilliant pressing. This scope showed no after-effects of forecasting the newness of food rather biosensors are utilized to make customary food stuffed more developed. It utilizes senior sensor-based way to deal

with tracking down the nature of the fluid sort of food rather than non-fluid pressed fronds [24] [25] [26].

3. Research Problem

Detecting the freshness grading level of the apple fruit is a bit difficult due to its varying varieties. Therefore, there is a dire need to design a machine-learning approach involving the digital image processing steps to classify the input apple fruit images into their correct classes. This will be handy to have healthy and fresh apple fruits for eating for good health.

This research study is focused on the classification of apple fruit freshness. Therefore, how can we classify the input apple fruit freshness in this thesis different research questions have been developed. These research questions are:

3.1. Research Questions No. 1

What are the various approaches available for the classification of the apple fruits freshness?

3.2. Research Questions No. 2

How can the machine learning approaches be adopted to classify the apple freshness into one of the three classes, including fresh, semi-fresh, or rotten?

4. Proposed Solution

The main objective of this section is to discuss and elaborate more evolutionary and scientific assumptions used in this research. Further, the implementation of the K-means clustering algorithm and neural network-based framework to find the current freshness state of apple fruits is discussed. K-Means is used to achieve the automatic grading of apples in terms of freshness. Finally, some limitations of the proposed research study are also highlighted.

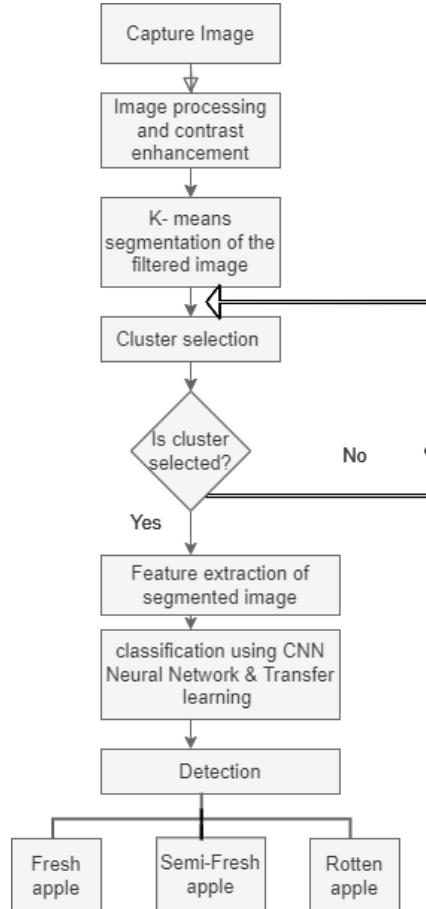


Figure 1: Model Flowchart

4.1. Proposed Methodology for Grading Apple Fruit Freshness

In this section, the suggested framework for packed apple fruit freshness classification is presented along with the relevant algorithms. The devised model is based on digital image processing using ANN and a well-known clustering approach, K-Means clustering. Figure 1 shows the fundamental working of the proposed approach.

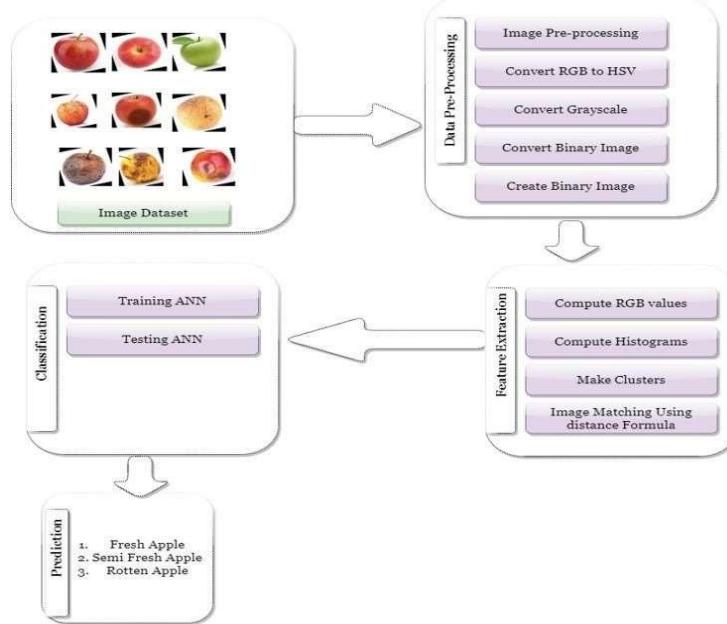


Figure 2: Proposed framework for the analysis of apple fruit freshness.

The image of the apple fruit is first segmented using K-means clustering, and then relevant features are extracted. The training set is then updated with these apple-extracted features. An ANN classifier is used for training, and the apples in the training set are labeled with the corresponding faults. An apple fruit test image is supplied for grading; the image is segmented and characteristics are extracted.

The corresponding class (defect) is then found by comparing the test apple image with the training images. The grading module classifies the test apple's grade by using the data on the indicated class (defect) and grading parameters. Figure 2 shows the freshness prediction algorithm.

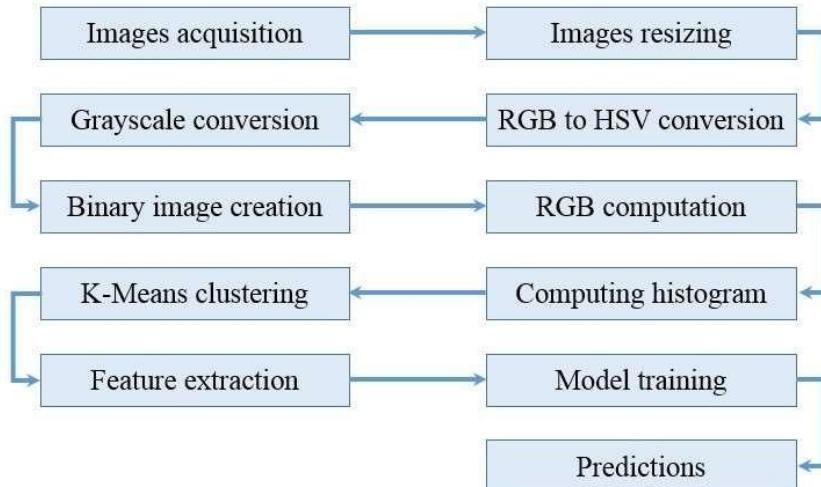


Figure 2: Packed apple fruit freshness prediction algorithm

4.2. Build Datasets

For experimentation, exactly the packed apple dataset could not be found. Therefore, for experimentation and performance analysis, a total of 756 raw images of apples are collected on three different days. Variations in time,

angle, and light intensity were purposefully added during the image-gathering process to improve the recognition model's accuracy and robustness. Using the same mobile camera (13 megapixels, F-Stop = f/2, exposure time = 1/50 s, ISO speed = 151, focal length = 4 mm, Redmi Note 4X, China), 345 original photos with a resolution of 3120×4160 pixels were taken:

Table 1. Table 1: The statistical details of the collected apple images

S. No.	No. of Days old	Class	No. of Images
1	20-30	Fresh	420
2	30-40	Semi Fresh	210
3	>48	Rotten	126

Data augmentation approaches have been used to overcome the problem of restricted training dataset size. When there is inadequate data, data augmentation offers a straightforward and practical way to improve model resilience, increase dataset diversity, and reduce over fitting problems

Table 2: Dataset composition after image augmentation on the training set.

Class	Size of Original datasets	Size of augmented datasets
Fresh	336	960
Semi Fresh	168	595
Rotten	100	575

4.3. Preprocessing of Datasets

Preprocessing the dataset is an important next step after compiling the packed apple dataset. To improve the sharpness, clarity, texture, and masking of the packed apple picture data, image preprocessing is utilized.

- Grayscale Conversion: In order to minimize computational complexity, the RGB images are converted to grayscale.
- Binary Image Creation: In order to streamline the analysis and highlight the key elements, the grayscale images are further transformed into binary images.
- Noise reduction: To ensure that the analysis is unaffected by unimportant variations in the image, an average filter is applied to remove Gaussian noise from the images.

4.4. Feature Extraction

Feature extraction is the process of using particular techniques to determine an image's key attributes. In this study, color and texture, two important characteristics, are used to evaluate the freshness of packed apple fruits. Features like as color, texture, and form are frequently used to assess the freshness of apple fruit; however, the success of these techniques mostly depends on feature extraction.

4.1.1. Color Pattern based on histogram

It is frequently recognized that fruit and vegetable color plays a critical role in quality recognition systems and customer satisfaction. The RGB color model is utilized in this study to represent color. Three discrete colors, red, green, and blue, defined in a 3_3 matrix make up an RGB image. Every image in the database has its RGB value computed and saved for comparison with query images.

$$H = (\cos)^{-1} = \frac{1/2[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \quad (1)$$

4.1.2. Morphology Processing

The analysis of structure is referred to as morphology. Mathematical morphology is used in medical image processing to recognize and extract significant picture features from the image based on its shape attributes. Morphological operations are logical transformations that compare the neighborhood of a pixel with a structural element, which is a predefined pattern [24].

Erosion and dilation are the two basic morphological processes. Dilation gives items the ability to enlarge, possibly

bridging tiny spaces and joining disparate elements. Conversely, erosion causes objects to contract by weakening their borders. More complex transformations can be produced by combining these two basic processes, erosion and dilation. Morphological filtering frequently uses opening and shutting processes. The same structural member is used for both dilatation and erosion during opening. On the other hand, closing involves dilation and erosion, essentially carrying out a reverse-opening process using the same structural element [25] [27].

K-Means clustering uses a clustering technique to perform feature extraction. It is an unsupervised learning algorithm, we used unlabeled data to define the different states of apple. K-Means divides the object into clusters that share the similarity and dis-similarity related to the object. A typical K-Means clustering algorithm is given in Algorithm 1.

Algorithm 1: K-Means clustering

- 1: Select K points as initial centroids
- 2: Repeat
- 3: Form K clusters by assigning each point to its closest centroid.
- 4: Recomputed the centroid of each cluster until the centroid does not change

There is a centroid point in an object, which is computed by minimizing the following objectives

$$k = q (X_o - X_C)^2 + (Y_o - Y_C)^2 \quad (2)$$

Where K denotes no of clusters, O is an observed value, and C is the centroid value, we performed this calculation to find the minimum distance of each cluster.

$$U = (X_o + X_n) / 2; Y_o + Y_n / 2 \quad (3)$$

Where U is new centroid value, Xo is an old cluster coordinate and Xn is new cluster coordinate, similar Yo and Yn.

4.5. K-Means Features Extraction

The main aim of K-means is to enhance the segmented image, so it will help an accurate classification of an apple if there is some defect. In this proposed approach, a square Euclidean distance is used to measure K-Means clustering. The algorithm classified and predicted their quality. The modified K-Means with square Euclidean distance for clustering for packed apple fruit freshness prediction is given in Algorithm 2.

Algorithm 2: Modified K-Means clustering

- Apply K-Means to sort out the colors of apples using the Euclidean distance method.
- Label each pixel in the image using the results of K-Means.
- Create color-segmented pictures from an input image.
- The median filter is then applied to the segmented image.
- Now the images are the input values for the classification procedure.

Figure: 5a.



Figure: 5b

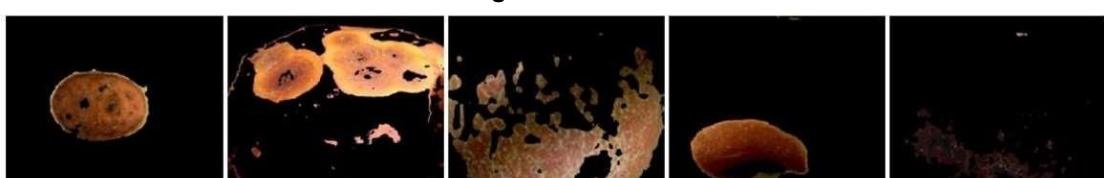


Figure 3: mage segmentation results, (a) Before segmentation, and (b) After segmentation.

5. Working of Proposed Approach

Image segmentation using the K-Means algorithm is very useful. The basic aim of image segmentation is to clear the background and object. Figure 5 shows the images without segmentation. Image segmentation helps in obtaining those areas that are affected, as shown in Figure 5b.

The main goal of the proposed approach is to segment color automatically using K-Means clustering and L*a*b* color space. The proposed framework works in the following steps:

- Import the image of the faulty fruits.
- Transform the RGB image into Lab* color space. Due to the presence of two chromaticity layers in the 'a*' and 'b*' channels and a luminosity layer in the 'L*' channel, the Lab* color space was selected. Because all color information is contained within the 'a*' and 'b*' layers, this selection is computationally efficient.
- To identify colors, use K-Means clustering in the 'ab' space. The difference between two colors is measured using the Euclidean distance metric.
- Using the K-Means clustering results, give each pixel in the image a label. An index for each cluster is assigned to each pixel.
- Create segmented pictures by using color segmentation as a basis. Using pixel labels to represent the amount of clusters found, the image is divided into many images.

Figure 6 shows the graphical representation of the defected clusters identification from a given input defected apple image.

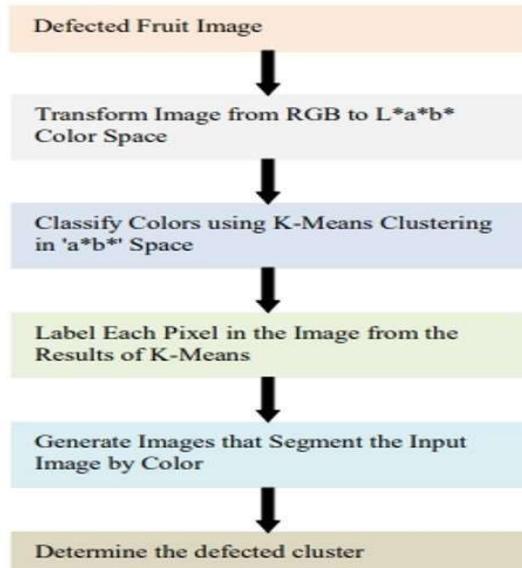


Figure 4: Discovery Process of defected clusters in defected apple fruit image

The algorithmic index of each cluster containing the faulty fruit piece is calculated, as K-Means may not always produce the same cluster index number. Alternatively, the cluster center value, which comprises the average value of 'a*' and 'b*' for each cluster, can be used to achieve this. Several images of damaged apples in packaging are displayed in Figure 7.

1.1. 5.1 Classification Using CNN Neural Network and Transfer Learning

In the suggested approach, the classification of the apple fruit freshness has been performed using the well-known and adopted transfer learning technique. Further, in the suggested transfer learning this study involved the adoption of the Alex Net architecture. In the subsequent paragraph, a detailed discussion of the architecture is presented. In Alex Net architecture, there are a total of eight layers. These layers are decomposed into two classes: The first, there are five layers in the Alex Net, which are called convolutional layers; and layers in the second class are fully linked layers. These two classes of convolutional layers and linked layers form the architecture of the Alex Net.

From the dataset of apple fruit images, images are read as input to the first layer of the Alex Net, and the size of all the images is converted to 256_256. This way, the output of the first layer is used as an input for the next layer, using 256 convolutional filters using a resolution of 5_5. In the next layers, the third and fourth layers of the model,

384 convolutional filters were applied, by resolution of 3_3. After the task of the third and fourth layers is finished, layer 5 comes into action and here using 3_3 resolution total of 256 kernels were applied. In convolutional layers, maximum pooling with the resolution of 2_2 was used using the scheme of batch regularization. Further, the significance of the selection of the activation function plays a vital role in enhancing the accuracy of the classification task. Also, the selection of the appropriate function must ensure that the value of the gradient function must converge rapidly and that the infinity value of the selected activation function should not return a value of 0.

For this study, the rectified linear unit (ReLU) function is used as an activation function, which very efficiently tackles the issue of the gradient, as discussed above. ReLU function has been selected as an activation function as it is the simplest and needs no complex processing for computing its required value and consequently, the chosen model may get its training in a short duration. Transfer learning is a very famous deep learning approach, which involves the reuse of the pre-trained model to enhance the accuracy of the model in use.

6. Results and Discussion

This study performs experiments with a self-collected dataset. As the dataset size was small, data augmentation was performed to increase the number of training samples. To avoid data leakage problems, augmentation is only used for training data and the testing phase uses the originally collected images. Finally, 1,800 apple sub-images were obtained. These generated images were added to the dataset the dataset composition. The dataset is divided into two sets for training and testing the proposed model. The training and testing sets were independent and randomly sampled from the originally collected images. In the ratio of 0.8 to 0.2. The training set was then augmented to increase its size for proper training of the proposed model.

6.1 Win-Loss Ratios

In Table 3, win-loss ratios are depicted, which are computed using the defective images of the apple fruit found in the dataset.

Table 3: Data samples for win and loss ratios.

Class Name	Total Images	Defected Images	Win (%)	Loss (%)
Fresh	164	-	95.7	4.3
Semi Fresh	50	8	84	16
Rotten	164	7	95.7	4.3

6.2 Performance Analysis of Proposed Approach on Augmented Dataset

In this experiment, we selected the same number of images of the apple fruits from the testing split of the dataset and we examined how the proposed approach classifies each image of the apple fruit from various classes. Experimental results show that the suggested model prediction of the rotten class apple was high, and it achieved a prediction accuracy of 94.34%. Table 4 shows the classification results of the proposed approach for the 'fresh', 'semi fresh', and 'rotten' classes of apple fruit indicating that the 'semi fresh' has a lower classification accuracy of 90.47%. The lower accuracy of the 'semi fresh' class is due to varying marks of the decomposition process found on apples. The nature and size of the decomposition are different leading to clusters of different sizes and numbers, thereby making it difficult for the model to discriminate between the 'rotten' and 'semi fresh' classes. For the 'fresh' class, the accuracy is 96.15% and the model obtains a mean accuracy of 93.95% accuracy.

Table 4: Results from testing dataset.

Apple State	No. of Apples	Apple states correctly predicted (%)	
		No. of apple	% of apple
Fresh Apple	84	80	95.23%
Semi Fresh Apple	42	38	90.47%
Rotten Apple	26	25	96.15%
Average	152	60	93.95%

6.3 Performance Analysis of Proposed Approach on Original Dataset

In addition to the augmented dataset, experiments are also carried out using the original dataset. Table 5 shows the confusion matrix for the proposed approach on the original dataset, as well as, the values for different performance metrics. It is observed that using only the original dataset, the accuracy of the proposed CNN-TL model is high with an accuracy score of 0.957, 0.94, and 0.96 for the 'fresh', 'semi fresh', and 'rotten' classes, respectively. Overall, a mean accuracy score of 0.96 is obtained with the original dataset while precision, recall, and F1 scores are 0.94, 0.95, and 0.98. Close agreement between these metrics indicates a good fit of the model and shows that the model does not exhibit over fitting.

The proposed CNN-TL model shows 18 wrong predictions for the 'fresh' class while 402 predictions are correct leading to an accuracy score of 0.957. For the 'semi fresh' class, performance is even better with only 6 wrong predictions. While for the 'rotten' class, the model made 6 wrong predictions out of 126 total predictions.

Table 5: Performance analysis of proposed approach on the original dataset

Apple state	Predicted class			Performance metrics			
	Fresh	Semi Fresh	Rotten	Accuracy	Precision	Recall	F1-Score
Fresh	402	12	6	0.957	0.93	0.94	0.98
Semi Fresh	6	204	0	0.971	0.94	0.96	0.99
Rotten	2	4	120	0.952	0.96	0.97	0.98
Average				0.96	0.94	0.95	0.98

6.4 Comparative Analysis with Deep Learning Models

For further exploration, different transfer learning models have been implemented in this study including ResNet, Google Net, Alex Net, VGG, MobileNetV2, and InceptionV3. These models are reported in existing studies for their better performance on image classification tasks and are adopted for performance comparison with the proposed CNN-TL model. Table 6 shows the comparative results of CNN-TL with other deep learning models implemented in this study. Results indicate a better performance of CNN-TL with a testing accuracy score of 0.96. Among the deep learning models, AlexNet and InceptionV3 models perform well, each with a testing accuracy of 0.95.

Table 6: Comparative analysis of the proposed approach with existing deep learning approaches using augmented and original datasets

Approach	Original dataset		Augmented dataset	
	Training	Testing	Training	Testing
ResNet	0.72	0.73	0.77	0.77
GoogLENET	0.93	0.92	0.93	0.92
AlexNet	0.96	0.95	0.95	0.93
VGG	0.94	0.95	0.94	0.93
MobileNetV2	0.87	0.88	0.88	0.87
InceptionV3	0.95	0.95	0.95	0.93
Proposed CNN-TL	0.98	0.96	0.97	0.94

6.5 Performance Comparison

As depicted in Table 7, the proposed CNN-TL model is compared with existing models from the literature. The performance of the proposed CNN-TL architecture was evaluated using the accuracy of the original dataset. For comparative studies, the novel suggested work in this study outperforms the existing models and thus acquires

the accuracy of 96% on the original dataset and 94% on the augmented dataset. This clearly shows the supremacy of the suggested CNN-TL model.

Table 7: Comparative analysis of the proposed approach with existing approaches using the original dataset.

Approach	Accuracy	Precision	Recall	F1 score
Tasnuva et al. Mukhiddinov et al. (2022)	0.94	0.93	0.89	0.93
Mudaliar Li et al. (2021)	0.91	0.95	0.87	0.94
Miriti Miriti (2016)	0.93	0.95	0.88	0.93
Valentino et al. Dubey and Jalal (2012)	0.91	0.93	0.87	0.91
Proposed Approach	0.96	0.94	0.95	0.98

7. Conclusions and Future Direction

This study endeavors to automate the classification of canned apple fruit using a convolutional neural network with transfer learning and K-Means clustering. A detailed analysis of the various apple fruit freshness techniques has been performed to explore their major contributions and limitations. Based on the identified limitations, this study proposed a novel image processing and convolutional neural network-based approach to analyze and check the freshness of apple fruit using the collected dataset of the apple fruit. A dataset of packed apples is gathered using an X-ray luggage scanner. Experiments are performed using the original, as well as, the augmented dataset. Experimental results indicate a similar performance of the proposed CNN-TL model with 96% and 94% accuracy for both datasets, respectively. Finally, the comparative analysis of the proposed approach with the existing approaches available in the literature has been conducted showing better results of the proposed model.

Although the introduced model in this study has shown very promising results and achieved 96% accuracy on the original dataset, still requires further improvements in accuracy. The proposed approach has the potential to use deep learning and the Internet of Things, which is the topic of many current papers in the field. We have used the proposed approach for apple fruit freshness classification; however, the described approach can be extended easily to classify other fruits including oranges, mangoes, peaches, etc.

8. References

- [1] S. R. Abdi, L. Chaves, and A. G. Jaeger, "An exploration of what freshness in fruit means to consumers," *Food Research International*, vol. 112491, p. 165, 2023.
- [2] R. N. S. and M. R. Pandey, "Image processing and machine learning for automated fruit grading system: A Technical review," *International Journal of Computer Applications*, vol. 81, pp. 29–39, 2012.
- [3] M. M. A. and C. J. Mukhiddinov, "Improved classification approach for fruits and vegetables freshness based on deep learning," *Sensors*, vol. 22, p. 8192, 2022.
- [4] A. and B. A. Bhargava, "Fruits and vegetables quality evaluation using computer vision: A review," *Journal of King Saud University-Computer and Information Sciences*, no. 33, pp. 243–257, 2021.
- [5] Z. Rashid, M. R. Khan, R. Mubeen, A. Hassan, F. Saeed, and M. Afzaal, "Exploring the effect of cinnamon essential oil to enhance the stability and safety of fresh apples," *Journal of Food Processing and Preservation*, vol. 44(12), p. e14926, 2020.
- [6] Q. Fu and N. Yao, "Grading methods for fruit freshness based on deep learning," *SN Computer Science*, vol. 3, p. 264, 2022.
- [7] P. V. V. K. Devi, et al., "Machine vision applications to locate fruits, detect defects and remove noise: A review," *Rasayan Journal of Chemistry*, vol. 7, pp. 104–113, 2014.
- [8] Y. N. M. and Y. W. Q. Fu, "Grading methods for fruit freshness based on deep learning," *SN Computer Science*, vol. 3, p. 264, 2022.
- [9] F. Boussaid, M. Kamel, N. Bouguila, C. Jiang, and D. E. Zhu, "Multiple hypotheses image segmentation and classification with application to dietary assessment," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 377–388, 2014.
- [10] J. A. M. and M. S. Dehais, "Food image segmentation for dietary assessment," *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pp. 23-28, 2016.

- [11] M. H. R. and B.-R. J. Buisman, "Discounting and dynamic shelf life to reduce fresh food waste at retailers," *International Journal of Production Economics*, vol. 209, pp. 274–284, 2019.
- [12] P. S. S. and A.-M. R. Pouladzadeh, "Measuring calorie and nutrition from food image," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, pp. 1947–1956, 2014.
- [13] A. S. I. and H. G. E. Krizhevsky, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84-90, 2017.
- [14] A. D. D. A. G. B. U. and S. S. Awate, "Fruit disease detection using color, texture analysis and ANN," *International Conference on Green Computing and Internet of Things (ICGCloT) (IEEE)*, pp. 970-975, 2015.
- [15] Y. F. X. L. Y. and H. X. Li, "Apple quality identification and classification by image processing based on convolutional neural networks," *Scientific Reports*, vol. 11, p. 16618, 2021.
- [16] S. R. and J. A. S. Dubey, "Detection and classification of apple fruit diseases using complete local binary patterns," *Third International Conference on Computer and Communication Technology (IEEE)*, pp. 346–351, 2012.
- [17] M. S. Miah, T. Tasnuva, M. Islam, M. Keya, M. R. Rahman, and S. A. Hossain, "An advanced method of identification of fresh and rotten fruits using different convolutional neural networks," in *Proc. of the 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1-7, July 2021.
- [18] G. ElMasry, N. Mandour, S. Al-Rejaie, E. Belin, and D. Rousseau, "Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview," *Sensors*, vol. 19, no. 5, pp. 1090, May 2019.
- [19] J. P. K. and L. R. Steinbrenner, "Hyperspectral fruit and vegetable classification using convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 162, pp. 364–372, 2019.
- [20] T. D. F. and L. M. Nordey, "Predictions of fruit shelf life and quality after ripening: Are quality traits measured at harvest reliable indicators?" *Postharvest Biology and Technology*, vol. 153, pp. 52-60, 2019.
- [21] S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis, and J. Blasco, "Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables," *Food Bioprocess Technol.*, vol. 4, pp. 487-504, 2011.
- [22] H. J. Heijmans, "Construction of self-dual morphological operators and modifications of the median," *Proceedings of 1st International Conference on Image Processing (IEEE)*, vol. 2, pp. 492-496, 1994.
- [23] A. M. Raid, W. Khedr, M. El-dosuky, and M. Aoud, "Image restoration based on morphological operations," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 4, no. 3, pp. 9-21, 2014.
- [24] K. B. D. M. and N. E. Stoknes, "Anaerobically digested food waste in compost for Agaricus bisporus and Agaricus subrufescens and its effect on mushroom productivity," *Journal of the Science of Food and Agriculture*, vol. 93, pp. 2188–2200, 2013.
- [25] R. and C. M. Tiwari, "Apple fruit disease detection and classification using k-means clustering method," *Advances in Intelligent Computing and Communication: Proceedings of ICAC (Springer)*, pp. 71–84, 2021.
- [26] S. Jamil, *The Challenge of Establishing World-Class Universities*, Washington DC: The World Bank, ISBN 978-0-8213-7865-6, 2009.
- [27] T. O. Writer, "THE Ranking Methodology 2016," 2016. [Online]. Available: <https://www.timeshighereducation.com/news/ranking-methodology-2016>. [Accessed 10 Aug 2016].
- [28] P. Haddawy, S. U. Hassan, C. W. Abbey, and L. I. Beng, "Uncovering fine-grained research excellence: The Global Research Benchmarking System," *Journal of Informetrics*, vol. 11, pp. 389–406, 2017.
- [29] Q. S. Writer, "QS World University Rankings Methodology," 11 Sep 2015. [Online]. Available: <http://www.topuniversities.com/university-rankings-articles/world-university-rankings/qs-world-university-rankings-methodology?page=1>. [Accessed 25 Mar 2016].
- [30] M. M. A. Chaudhry, M. L. Amodio, F. Babellahi, M. L. V. de Chiara, J. M. Amigo Rubio, and G. Colelli, "Hyperspectral imaging and multivariate accelerated shelf life testing (MASLT) approach for determining shelf life of rocket leaves," *J. Food Eng.*, vol. 238, pp. 122-133, 2018



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:

Ujan et al. (2024). "Proposed Model of Cloud Health Information Systems in Pakistan". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Proposed Model of Cloud Health Information Systems in Pakistan

Dr. Imran Anwar Ujan^{a*}, Asadullah Shah^b Dr. Noor Afzan Salleh^b

^a Department of Information Technology, Faculty of Engineering & Technology, University of Sindh, Jamshoro (uijanium@gmail.com).

^b Kulliyah of Information & Communication Technology, International Islamic University, Malaysia. (asadullah@iium.edu.my)

* Corresponding author: ujanium@gmail.com

Abstract:

Cloud computing, a cutting-edge improvement in Information Technology, offers an alternative method for managing and obtaining health data. This technology is dependent on several web-based software applications that are accessed by a large number of users over the internet. Pakistan's healthcare sector is now in the nascent phase of integrating cloud computing technology. However, the amount, variety, and speed at which healthcare data is growing in both government and nongovernment hospitals in this country is increasing significantly. Therefore, to effectively store and handle healthcare data, the medical industry in Pakistan should recognize the significance of cloud computing. The survey will analyze several elements that are influencing medical specialists at certain hospitals in Pakistan to embrace cloud-based health information systems. This study utilized a model to propose that the adoption of CHIS is influenced by 13 primary constructs. These constructs include Behavioral Intention (BI), Perceived Usefulness (PU), Usability (UB), Cost Effectiveness (CE), Internet Work (IN), Facilitating Conditions (FC), Performance Expectancy (PE), Compatibility (COM), Complexity (CMP), Data Security (DS), Data Privacy (DP), Hardware Modularity (HM), and Software Modularity (SM).

Keywords: Cloud Computing ; Health Information Systems; Pakistan.

1. Introduction

The field of information technology has experienced a significant and quick expansion since the start of the 21st century, resulting in substantial impacts on information systems and associated domains. It is important to recognize that health information management is not an exception in any way. Information technology has played a pivotal role in the advancement of healthcare systems. As a result, it enhanced the quality of healthcare and improved access to healthcare services. It led to a clear decrease in both medical expenses and mistakes. However, it is crucial to give careful consideration to IT adoption while implementing any Health Information Systems (HIS). As per the Medical Records Institute (2003), healthcare organizations want to achieve internet-based health applications such electronic health records, electronic prescription, and mobile health. Various studies have addressed the topic of using the Internet to improve the efficiency of the healthcare industry and minimize errors in care delivery operations. However, health care organizations are currently experiencing a delay in the adoption and use of health information systems, despite their best efforts (Abdekhoda; Ahmadi, Dehnad;

Hosseini, 2014). The rapid advancements in cloud services have led to numerous consequences in the dissemination of healthcare. The latest electronic health systems continue to face challenges including internet access, expenses, customer assistance, and disaster recovery (Laupacis et al. 1992). The medical business can greatly benefit from the deployment of cloud computing in this context.

Cloud computing services, which are considered a service solution and rely on cloud processing, are responsible for the administration and processing of health-related data in a distributed health environment (Philipson & Jena 2013). In the present day, cloud services offer a versatile solution for improving the efficacy and competitiveness of an institution. Contemporary technology offers innovative delivery methods; however, the processing and application of services are essential. The low application of cloud services is not completely comprehended by emerging countries (Sant'Anna et al. 2007). This may be due to the dynamic configuration of cloud computing, which provides computing services and a variety of operating capabilities. These capabilities are primarily used for the data administration of an organization in a widespread, universal, and prevalent manner. Software, systems, and platforms are supported by cloud computing in autonomous locations. Additionally, it consists of services that are accessible from any location via the Internet and are shared (Schuktz & Wanda 2004).

2. Significance of the Research

If we want to rebuild and reform our health care system, we need health information systems and new forms of communication and information technology. The value of a health information system to a hospital can be better



Published by
Foundation University
Islamabad.

Web: <https://fui.edu.pk/>

location via the Internet and are shared (Schuktz & Wanda 2004).

understood by upper management. Patients in underdeveloped nations will have access to top-notch medical care in a timely manner once the HIS is implemented in hospitals there.

Academically, the current study will contribute to the research on the acceptability and utilization of the HIS by expanding the Technology Acceptability Model (TAM) and D&M IS success model. The objective of the present investigation is to offer an alternative viewpoint. This will be a substantial contribution to the literature in both a theoretical and practical sense, as it will identify the variables that facilitate the adoption of the HIS. The significance of utilizing HIS applications in clinical practice is acknowledged in this study. The objective of this investigation is to substantiate the significance of medical practitioners' training and education in order to enhance their adoption of the HIS as a standard in the clinical judgment process. The objective of this investigation is to encourage the implementation of programs that enhance the IT capabilities of healthcare professionals, and, as a result, the establishment of training facilities that will enhance the information technology capabilities of physicians. Additionally, the objective of this investigation is to underscore the necessity of computer literacy and information technology training programs for physicians in order to advance evidence-based practice. An increase in the productivity and affectivity of evidence-based practice will be the long-term consequence of prioritizing the development of computer and information skills. The significance of skills infrastructure is much greater than that of physical infrastructure, as demonstrated in the literature.

This study provides a concise framework for addressing the determinants of health technology, acceptance, and user intention toward digital clinical and medical information resources, making a theoretical addition to the field of health informatics in general.

3. Proposed Research Framework

Based on the existing literature in information system research, researchers continue to face a recurring challenge of choosing the most suitable model from a range of options and determining the implementation of new technologies in practical settings (Karahanna and Straub, 1999; Moore and Benbasat, 1991). In order to anticipate the factors that influence the acceptance of technology and enhance comprehension of this process, researchers have combined, adjusted, and utilized several theoretical frameworks from various fields such as social psychology, information systems, sociology, and business management (Karahanna and Straub, 1999).

Researchers face a challenge in selecting the most effective methods to achieve their objectives as a result of the abundance of information available. This is due to the fact that a model that was chosen, despite appearing to be appropriate for the research objectives, may not yield the desired results, even if previous studies have employed the same model in a similar study, rendering the remainder of the models useless (Chen et.al 2002). This complicates the comprehension of the causal pathways of relationships. Conversely, underflow transpires when the intended result fails to materialize. Ideas from multiple models could be consolidated to align with the modified model's objective, thereby preventing overflow and underflow (Karahanna, and Straub, 1999).

The measurement of success in information systems has been a subject of interest to numerous information systems researchers. The majority of research has focused on the evaluation of information systems' efficacy. The study conducted by D & M (1992) has been the most influential in terms of the efficacy of IS. Information systems success is recognized as a multidimensional phenomenon, which is classified into six categories: 1) information quality, 2) system quality, 3) user satisfaction, 4) use, 5) organizational impact, and 6) individual impact. To develop an inclusive instrument for measuring IS effectiveness, researchers should consider contingent factors such as individual, organizational, technology, and task factors, as D&M (1992) argues that it is questionable to measure these six categories alone.

Alternatively, Myers, Kappelman, and Prybutok (1997) proposed an inclusive IS assessment framework based on the integrated work of Saunders and Jones (1992) and D&M (1992), incorporating organizational and external environmental factors into the model. Much research has combined the D&M (1992) model method with other contingent elements such as individual, organizational, technological, and task aspects (Jen and Chao 2008, Lee and Chung 2009, Molla and Licker 2001, Zhang et al. 2005). Ballantine et al. (1998) extend this model to create a three-dimensional IS success model. Ballantine et al. (1998) categorize IS success into three levels: deployment, development, and delivery. However, because of its complexity, the model proposed by Ballantine et al. (1998) has not been empirically evaluated.

Previous research on IS success incorporates the D&M model as well; it also classifies IS success and recognizes the varying perspectives of stakeholders on success. As a result, it is deemed appropriate for both theoretical and empirical research, and it has gained widespread acceptance (Ballantine et al. 1998; Seddon et al. 1999). The IS success model as viewed by D&M (1992, 2003) is integrated into the theoretical framework. The goal of this study is to create an implementation framework for HIS and a D&M IS success model that will probably match the framework given that HIS is a form of IS.

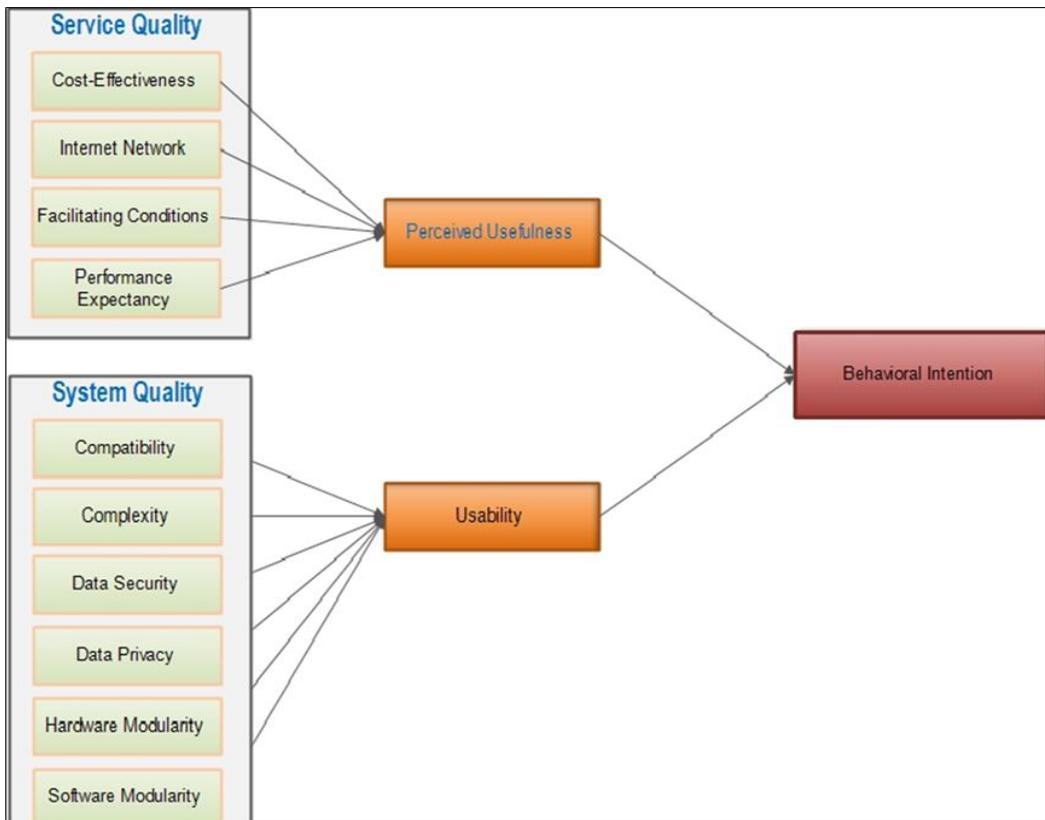


Figure 1: Impact of Service and System Quality on Behavioral Intention

4. The Research Methodology

The contemporary integration of technology in numerous industrialized nations has transformed the management and exchange of data by healthcare professionals, the utilization of that information by diverse entities, and the approach to addressing health-related issues. The integration of contemporary technology, such as cloud computing, inside Pakistan's medical industry is currently at a complex phase, and its effective use could yield collaborative and innovative platforms for healthcare information management. Technology integration in the medical data system concerning cloud services remains questionable. Technicians and clinicians utilize diverse specialized technical platforms to enhance the quality of health services. This research aims to model and emphasize the interrelationship among the key components for the optimal utilization of health data services dependent on cloud computing by both governmental and non-governmental health sectors in Pakistan.

A cross-sectional survey will be used to collect data for this research study because it is the most reliable method, doesn't cost much, and can be used to get a pretty large group size. Major hospitals in Karachi, Lahore, and Islamabad that already use the HIS system will be asked to give the information. The method is already being used by health care workers in these hospitals.

The questionnaire will be based on the conceptual model. The seven-point Likert scale will be utilized throughout the questionnaire. The field survey will be done following the pre-testing of the data collection instrument. The valid responses will be coded and analyzed using SPSS version 25 and AMOS. Both descriptive and inferential analysis will be used.

5. Conclusion

This study examines the potential effect of certain factors on healthcare professionals' utilization of cloud systems in Pakistan's public and private hospitals. The results will be obtained from examining the relations between certain service structures and system-related factors. The individuals' behavioural factors will be found to support all the proposed hypothesis. It is also anticipated that the outcome of the research will provide the roadmap for healthcare policymakers in Pakistan to how advanced technology will be utilized in the healthcare sector.

6. References

- [1] Abdekhoda M, Ahmadi M, Dehnad A, Hosseini AF. Information technology acceptance in health information management. *Methods of information in medicine*. 2014 Jan 1;53(1):14-20.
- [2] Karahanna, E., and Straub, D.W.: 'The psychological origins of perceived usefulness and ease-of-use', *Information & Management*, 1999, 35, (4), pp. 237-250.
- [3] Laupacis, A., Feeny,D., Detsky, A,S, & Peter, X,T,1992. How attractive does a new technology have to be to warrant and utilization? Tentative guidelines for using clinical and economic evaluations. *CMAJ: Canadian Medical Association Journal* 146(4); 473-481
- [4] Moore, G.C., and Benbasat, I.: 'Development of an instrument to measure the perceptions of adopting an information technology innovation', *Information systems research*, 1991, 2, (3), pp. 192-222.
- [5] Schultze, U. & Wanda, J.O.2004. A practice perspective on technology-mediated network relation: The use of Internet-based self-service technologies. *Information system Research* 15(i): 87-106.

An Intelligent Neural Network Based Approach for Enhancing Text-Based Information Retrieval Performance Using Social Networks

FICSE

Research Article

Article Citation:

Ahmad et al. (2024). "An Intelligent Neural Network Based Approach for Enhancing Text-Based Information Retrieval Performance Using Social Networks". *FUSST International Conference on Computer Science and Engineering*

Tauqeer Ahmad^{a*} Asif Nawaz^b Muhammad Rizwan Rashid Rana^c

^a University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan.
(tauqeer.ahmed443@gmail.com).

^b University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan.
(asif.nawaz@uar.edu.pk).

^c University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan.
(rizwanrana315@gmail.com).

* Corresponding author: tauqeer.ahmed443@gmail.com

Abstract:

Twitter is performing very important role in the field of social media to retrieve the interests of people. The tweets contain various messages that garner the highest number of retweets and likes, reflecting the demand and interest of the users. By sharing informative content, users can enhance the popularity of their tweets. Additionally, indicating the number of retweets and likes before posting a text online can contribute to increasing the user's post reputation, particularly when the content is of high quality. In the assessment of information retrieval, social media has a significant role in the evaluation of various aspects. This project demonstrates how information is gained from social media in the interest of the user, this information is then used to augment the user's base data, the priority is on the content of the tweets, which should lead to a large number of followers and a sufficient number of retweets. The top quality of tweets is beneficial to the online reputation of people who have a passion for different fields. An intelligent neural network that recognizes the most relevant keyword for a particular period of time and situation that is associated with web content in a temporal manner. The proposed model's primary advantage is that it effectively finds the information the user is looking for. If a user tweets and likes the text prior to posting it on a social media platform, the result will demonstrate that the proposed models are effective with the Intelligent Neural Network algorithm, which produces more effective and superior results.

Keywords: Text Mining; Twitter; Information retrieval; Social context; Social information retrieval; Query expansion Page ranking, Classification, NN algorithm.

1. Introduction

Information retrieval is a communication procedure that involves the exchange of information between the user and the server via an interface. As a result, users can access documents, visualizations, sound recordings, and records that satisfy the requirements or goals of the connected individuals of an information system or service [1]. This style of communication is characterized by documents that are intended to be accessed later, but are instead created as messages that are observed as documents, before they are actually used. Instead, they are placed in some form of storage, and then accessed by people who have a simple query that can easily be performed years after their creation. Users understand the knowledge and expertise associated with it and know how to most effectively utilize it in their favor. Its outcome is not always obvious how to most consistently implement this model.

Copyright
Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

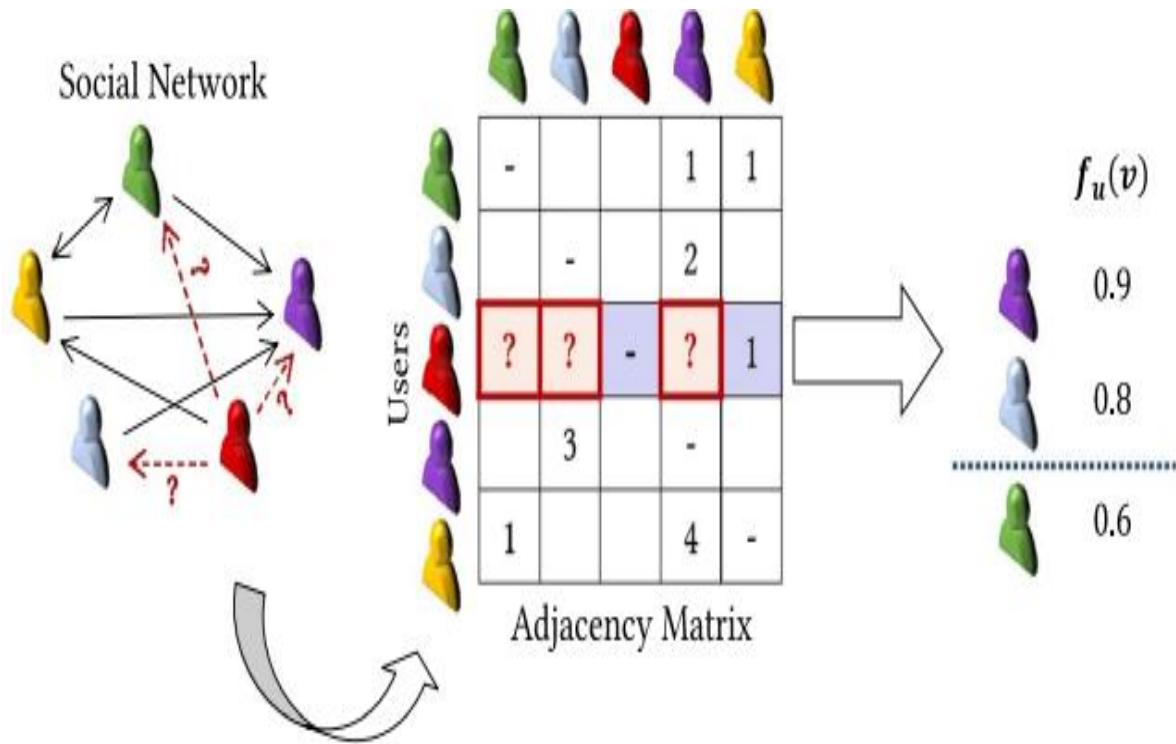


Figure 1. The SNS recommendation Matrix.

Social media platforms come in various forms, including Facebook blogs, business forums, podcasts, microblogs, photo sharing sites, product/service reviews, and weblogs [2]. To analyze these platforms, semantic networks have been employed to explore the underlying meaning of the text [3]. This is achieved using edge-labeled directed graphs, where concepts are depicted as vertices and their relationships as edges. A semantic network, a form of predicate logic, gathers and represents data. Consequently, the relationships within domain ontology concepts, assessed for document relevancy, have been utilized [4].

Researchers have recently discovered that social networks are valuable for tracking the evolution of various aspects of information retrieval (IR). Short texts on these platforms often have sparse feature information and limited semantic representation [5], [6], [7], [8]. Social Information Research (SIR) models represent a new generation of IR systems that leverage social networking content as a secondary information source to enhance social networks and improve the quality of research results [9].

The principle of "wisdom of the crowds," defined by the author, refers to the collective cognitive capacity of Internet users who contribute to web resources through comments, annotations, and opinions [10], [11]. The vast amount of user-generated data brings to light major IR challenges, such as the relevance calculation systems and the efficient use of this data in the IR process [12].

Traditional IR systems can calculate a resource's relevance based on its content parameters, but the current challenge is integrating diverse social data into the relevance calculation model. A new metric has been proposed to address this challenge [13]. Compared to state-of-the-art approaches, this new social data analysis method has a significant impact. Many businesses and individuals use social media platforms like Twitter, WhatsApp, and Facebook for advertising, selling, and purchasing, especially among younger generations [14].

To achieve successful information retrieval in social networks, a new Intelligent Neural Network-based Information Retrieval (NNIR) technique has been introduced [15-17]. This study demonstrates that groups formed based on document relevance offer superior classification accuracy compared to previous models. Building upon this foundation, the research proposes a novel Intelligent Ontology and Neural Network-based Information Retrieval (IONN-IR) model, specifically designed for effective information retrieval within social networks. The IONN-IR model leverages ontology-based classification to enhance semantic understanding of the content, thereby improving document relevance grouping. By integrating neural networks, the system enhances pattern recognition, enabling the automatic detection of complex relationships between social network content and user queries.

1.1 Research Contribution

This study focused on retrieving user interest-based information from large social media datasets, specifically tweets and retweets. The research aims to make three key contributions:

- i. We provide a practical solution to assist media monitoring staff in mining texts and tweets about specific themes, thereby enhancing media monitoring analysis reports based on text semantics.
- ii. We demonstrate a knowledge discovery strategy using an intelligent neural network (NN) and page ranking from a selected cluster of interesting events tailored to user interests.
- iii. By improving the syntactical quality and resolving problem complexity, we enhance the effectiveness of user query communication, leading to increased accuracy and efficiency.

The rest of the paper is organized as follows: Section 2 covers related work, Section 3 describes the proposed model, Section 4 presents experiments and evaluation, Section 5 discusses the experimental results, and Section 6 concludes with future work.

2. Related Work

In this section, we discussed various recent research studies, focusing on their methodologies, algorithms, and technical insights. Currently, Twitter plays a significant role in social media, reflecting human behaviors through user interests [18]. Users' retweets and likes on posts can indicate their reactions before sharing content on social media. This paper employs several machine learning classifiers to predict the number of likes and retweets a post will receive. Our results demonstrate that using the bag-of-words technique improves model performance by 10-15% compared to older methods.

Extensive efforts have been made to enhance the accuracy and diversity of tag-based image retrieval [19]. The author introduces a novel method to diversify retrieval results through latent topic analysis, starting with an initial relevance score estimation using Non-Negative Matrix Factorization (NMF) [20]. Efficient text retrieval from large datasets has become a significant topic in the information retrieval (IR) community, especially concerning users' privacy issues [22].

Today, digital text connects all types of professional, educational, and other activities. Digital instruction, success parameters, and materials solve real-world problems more effectively [22]. The author examined both structured and unstructured data on the internet, addressing the challenges of effective information retrieval posed by the vast amount of data online [23]. This technique reduces dimensional complexity in the feature vector, enhancing information retrieval.

A novel approach using an effective algorithm has been developed for contact recommendations in social media networks. Identifying beneficial user connections in a network has traditionally been challenging [24]. Further improvements in effectiveness were investigated using IR models as neighbor selection methods with user-based and item-based (KNN) collaborative filtering approaches.

A significant amount of small text data has emerged on the web, characterized by simplicity, immediacy, interactivity, and honesty [25]. To address these issues, new limited text retrieval methods have been proposed, leveraging the latest semantic knowledge sources like Wikipedia [26]. A novel semi-explicit small text retrieval method is suggested, combining Wikipedia features and related topics to sort and compute semantic similarity between target short texts and user queries [27].

Table 1. Literature Review Comparison Table

Paper Name	Problem	Technique	Results	Limitation
Selvalakshmi et al., 2019 [28]	relevancy issues were big issue.	Syntactic approach, page ranking algorithms.	Information Retrieval Time Reduction. Improve the retrieval accuracy and Relevancy.	User queries provide sufficient communication for effective distributed processing.
Wu et al., 2020 [29]	The protection of user privacy is becoming an increasingly significant.	A client-based framework and privacy model were utilized.	Dummy sequences were constructed to protect privacy, resulting in improved performance.	Efficiency and effectiveness are very low, with issues in space allocation and time overhead.
Kumar et al., 2020 [30]	1. Complexity increases with a greater number of features. 2. Time complexity.	Random forests, Naïve Bayes, SVM and Decision tree.	Reduce the dimensionality of the vector space while maintaining a high level of accuracy and improved performance.	Accuracy issues arise, a small dataset is used, and architecture complexity has increased.
Sanz-Cruzado et al., 2020 [31]	Compelling problems arise when people interact in online networks.	KNN , M25 and VSM	The information retrieval model proves effective in neighbor selectors, samplers, and features in learning to rank.	Enable better results in search tasks, confirming the need for improvements.
Li et al., 2020 [32]	Short text has a limited semantic representation, and the process was not effective.	1. Feature selection involved the use of the ESA model and LDA model.	Achieving better performance than state-of-the-art systems.	Short text classification

3. Propose Model of Information Retrieval using Neural Network Approach

In this section, we discussed various social media datasets that users can access based on their interests. The selected datasets can be retrieved and utilized in the user interface module for information retrieval using an intelligent neural network model. This model clusters the chosen documents within the dataset into different groups. After clustering, a page ranking process is applied according to specific rules. An intelligent, information retrieval-based neural network technique is then used to achieve improved results.

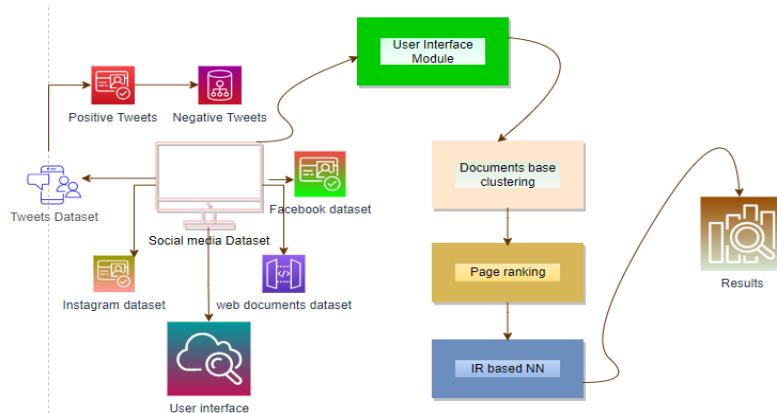


Figure2. Proposed Methodology Diagram

In the proposed methodology diagram, various social media datasets are utilized, allowing users to retrieve datasets based on their interests. The selected datasets can be accessed and used in the user interface module for information retrieval via an intelligent neural network model. This model includes a clustering component designed to organize the document-based datasets into different clusters. Following clustering, a page ranking process is applied using specific rules. An intelligent, information retrieval-based neural network technique is then employed to achieve enhanced results.

3.1. Data Set Used

The most frequently asked question is what dataset is selected. The sections below provide quick answers to this question.

3.2. Twitter dataset

A large volume of Twitter datasets was obtained from the relevant dataset repository based on public interest for utilization and privacy protection of users. Despite this, Twitter datasets remain physically accessible to users. This API uses Open Authorization to authenticate users and Twitter apps, delivering responses in JSON format [32]. The dataset comprises a corpus of tweets, which included numerous errors, emoticons, typographical mistakes, and unusual characters, necessitating extensive preprocessing.

3.3. Web documents

Huge of the web document datasets retrieved on public interest from the web directory. Some of these web document datasets identified physically. So, it was very easy for user choose to one of relevant data that was taken from the web document REST API; by using the REST API, and user can able to programmatically retrieved web document data [32].

3.4. Facebook dataset

Mostly Facebook datasets users used for further processing, it came from public priority requests, which can be retrieved for relevant tasks with Facebook's rules for data retrieving process, the user can use for privacy disclosure. So, there are different API used in Facebook to retrieve data set on public application [32]. In the latest study, it is observed that more than 20,000 Facebook comments were randomly selected for analysis. However, extensive preprocessing was required for this dataset due to numerous errors, emoticons, typos, and peculiar characters present within it.

3.5. Data Pre-processing Algorithm

In this section, the following steps performed for the preprocessing of the data set which are;

Step 1: Clean Tweets Using Regular Expressions

```
tweets <- gsub ("http\S+|www\S+", "", tweets)
tweets <- gsub ("@\S+", "", tweets)
tweets <- gsub ("RT", "", tweets)
tweets <- gsub ("[\x01-\x7F]", "", tweets)
```

Convert cleaned tweets to Python list for further processing.

```
tweets = tweets. to list ()
```

Step 2: Remove Stop Words Using NLTK

Common words like "and," "are," and "this" are widely used, but not for knowledge sources. It is important to delete from textual data.so, we use algorithm to remove unnecessary data by using NLTK tool.

```
import nltk
from nltk.corpus import stop_words
nltk.Download('stop words')
stop_words = set(stop_words.words('english'))
```

```
tweets = [" ".join ([word for word in tweet. Split () if word. Lower () not in stop words]) for tweet in tweets]
```

Step 3: Remove Duplicate Tweets

```
tweets = list(set(tweets))
```

Step 4: Part of Speech tagging/Information extraction

A tagging algorithm takes a sequence of words and a tag set as input and creates a set of tags with a single best tag for each word. This gives accurate data about the phrases and its context.

3.6. User Interface Module

Information retrieval model can enhance the performance by using User Interface Module (UIM) which transfer dataset on the interest of users with the help of using different social media sites such as twitter, Instagram, YouTube or Facebook. So, classification and clustering processes perform on selected data set by using UIM carrier.

3.7. Document Base Clustering

Document-based clustering is a creepy technique in information retrieval that involves grouping Divide documents into clusters based on content similarity. This approach improves efficiency and Efficiently retrieve relevant information by organizing large volumes of documents into a manageable form and meaningful clusters. By analyzing and classifying documents, clustering algorithms can identify patterns and relationships in the data, making it easier for users to find related documents. This process Helps reduce information overload, improve search accuracy, and provide increasingly coherent and consistent information Contextual search results.

3.8. Page Ranking

The evaluation and prioritization of tweets in information retrieval from social media datasets, such as Twitter, revolves around determining the significance and importance of tweets in relation to user preferences. This intricate process commences with the collection and preprocessing of data to ensure the text is purified and standardized. By scrutinizing past activities and interactions, user interest profiles are constructed through the extraction of essential elements like keywords, hashtags, and engagement metrics. Subsequently, each tweet is assessed and assigned a relevance score based on its alignment with these user interests. Modified versions of the PageRank algorithm are implemented to analyze the interconnectedness of tweets and user engagements, assigning greater significance to tweets that receive frequent interactions from influential users. The ultimate ranking is a fusion of relevance scores and PageRank outcomes, resulting in a personalized compilation of tweets that resonate with the user's preferences and generate substantial engagement. This process is consistently fine-tuned through user feedback, ensuring an optimized and tailored tweet selection.

3.9. Information Retrieval Base Neural Network

Utilizing deep learning techniques, neural networks revolutionize information retrieval (IR) by extracting meaningful information from extensive datasets. By employing models like CNNs, RNNs, and transformers such as BERT, neural networks have the ability to convert raw text into continuous vector spaces, effectively capturing semantic meanings. This surpasses the capabilities of traditional methods by automatically learning complex patterns and representations. To achieve this, these models are trained using vast labeled datasets to master relevance scoring, enabling them to rank documents based on their alignment with queries.

3.10. Results

The methodology integrates data from social media platforms like Twitter, Facebook, and Instagram, categorizing tweets by sentiment. The User Interface (UI) module captures user queries and preferences, directing them to the document clustering process which organizes the datasets into similar groups. Page ranking then prioritizes documents within each cluster based on relevance to the user's query. These ranked documents are further refined using neural network models for enhanced precision and contextual accuracy. The final, highly relevant results are presented through the UI, ensuring improved retrieval efficiency, enhance the relevancy, and user satisfaction by combining clustering, ranking, and advanced neural network techniques.

4. Results and discussion

An intelligent neural network model is proposed and experimented in enterprise information retrieval model, and compared with the results and performance of other existing models. The system was built by using the Java programming language. Users can retrieve datasets or information from social media sites based on their specific interest's area. Proposed model performance with various types of groups of media sites and different number of tweets in various clusters that can be retrieved using an intelligent neural network technique. A neural network (NN) technique is used to identify appropriate keywords that are semantically related to web articles. On the basis of user interest in today 's environment, the archived results show the significance and importance of the proposed technique. This technique produces more efficient results than the previous technique by employing a different classifier.

4.1 Data Collection

1	Number o	Number o	Number o	Number o	Reconstru	Insured di	Total ecor	Death rat	Injury rat	Number o	Homeless	Total num	Number o	Number o	Number o	Number o	F
2	0	0	0	0	0		0	0	0	0	0	210	200	0	0	0	200
3	0	0	4800	0	4800	0	0	20	0	0	44.06095	0	44.06095	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	6.1	1.5	9000	0	0	9001.5
5	0	0	0	0	0	0	0	0	0	0	0	51.3	351.8	6244	658	7253.8	
6	0	0	0	0	0	0	0	0	0	0	0	742.6	358.3	27168.5	7702.5	35229.3	
7	3.7	0	476000	0	476000	0	0	5	0.017806	0	2061.38	0	2061.38	123.6	151.2	9074	1569.5
8	0	0	1525000	0	1525000	0	0	14200	0	0	4212.833	0	4212.833	21.7	5830.5	339	3627.5
9	1100	0	0	0	0		0.832564	0	0	0	0	10	0	0	0	0	
10	8500	0	3200	0	3200		5.988296	0	2.254417	0	2.254417	1.2	0	0	0	0	
11	2400	0	0	0	0		1.544764	0	0	0	0	1.2	0	0	0	0	
12	0	0	0	0	0		0	0	0	0	0	2.2	13	0	750	763	
13	5000	0	0	0	0		2.436099	0	0	0	0	27.6	0	0	0	0	
14	0	0	0	0	0		0	0	0	0	0	128.3	502.8	0	12925	13427.8	
15	205	0	963348.3	0	963348.3	0	0	19570	0.064073	0	280.6868	0	280.6868	1257.7	2703.1	4650	1940.5
16	11900	0	1951200	0	1951200	0	0	40720	3.027666	0	452.6997	0	452.6997	0	0	0	0
17	55495.4	0	8256607	0	8256607	0	0	7275.4	10.81799	0	1581.907	0	1581.907	297	1015.3	51609.3	49320
18	23.9	0	8618425	0	8618425	0	0	224493.9	0.003023	0	1216.464	0	1216.464	79.9	1059	13045	4027
19	88.8	0	11948903	0	11948903	0	0	0	0.010273	0	1297.822	0	1297.822	333.5	1303.6	13461	23910.4
20	2000	0	12552200	0	12552200	0	0	365300	0.192437	0	1095.536	0	1095.536	3.1	89.2	15142.1	69
21	0	0	0	0	0		0	0	0	0	0	12	0	0	0	0	
22	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	
23	0	0	0	0	0		0	0	0	0	0	20	0	0	0	0	

Figure 3. shows Dataset of Natural disaster

In the Covid-19 and omicron millions of tweets of natural disaster retrieved from different social media cites by using different library like tweepy for information retrieval process. "Information Retrieval" is the very active domain both in industry and academia that achieved tremendous momentum in recent years. So, information retrieval process is selected for this study and tweets with different tags to extract data from the large database.

Table .2 Dataset

Features	Scale
Total number of tweets	45,109
Number of events	14
Total number of retweets	21,319

4.2 Analysis of Text Retrieval Results

This dataset comprises 22,900 documents, with a subset of short query documents reserved for testing purposes. It encompasses a total of 519,596 terms and user query annotations. Precision, recall, and F-measure are employed as metrics to evaluate the text retrieval performance of the neural network. These metrics are defined in Equations (1) and (2): To assess and produce model outcomes, a substantial Twitter dataset [34] was utilized.

$$Percision = \frac{X}{X + Y} \quad (1)$$

$$Recall = \frac{X}{X + Z} \quad (2)$$

$$F - Measure = \frac{2 * Percision * Recall}{Percision + Recall} \quad (3)$$

These terms like, x shows the total number of properly classified documents, y and z is highlighting the number of false positives and negatives accordingly.

Total record consider	LDA Technique results		SOR Technique results		Intelligent NN classifier	
Number of Features	Full	Selected	Full	Selected	Full	selected
1000	0.29	0.27	0.74	0.26	0.19	0.16
2000	0.38	0.32	0.66	0.27	0.22	0.20
3000	0.47	0.37	0.35	0.17	0.14	0.99
4000	0.55	0.42	0.23	0.37	0.34	0.29

Table 3 Comparative time analysis

In the above experiment, data retrieved from five different tweets were retrieved from the provided table. These datasets consist of varying numbers of records and were used for comparative time analysis. Existing techniques such as LDA and SOR were applied alongside a proposed technique, which demonstrated superior results compared to the analysis conducted with existing classifiers.

Table 4 Performance comparison between different techniques [35]

Total record consider	LDA Technique results		SOR Technique results		Intelligent NN classifier	
Number of Features	Full	Selected	Full	Selected	Full	selected
1000	91.24	91.26	92.34	92.36	99.12	99.14
2000	91.26	91.28	92.37	92.39	99.16	99.19
3000	91.31	91.33	92.39	92.41	99.19	99.22
4000	91.33	91.35	92.42	92.44	99.22	99.26

Relevant records taken from above five different tables also retrieved from table 6 which has different number of records of relevant Performance analysis and then previous techniques applied like LDA and SOR with current technique which shows better results as compared to the existing classifiers [36].

Features and characteristics, including selections like LDA [37] and SOR [38], are taken into consideration. The mean average precision (MAP) is determined by calculating the arithmetic mean of the average precision values across a set of n query topics [39].

$$\text{MAP} = \frac{1}{n} \sum m_a P_n \quad (4)$$

In this equation, AP denoted the precision value and MAP which is used metric for calculating the relevancy of the search results.so, we can identify the F-measure weighted harmonic mean of precision and recall.

$$F = 2 \frac{P \cdot R}{P + R} \quad (5)$$

Here, P indicates precision, which measures the retrieval of related documents. R signifies recall, which determines the number of relevant documents that can be retrieved from the provided search results. [40] introduces macro averaging and micro averaging metrics for evaluating the search results.

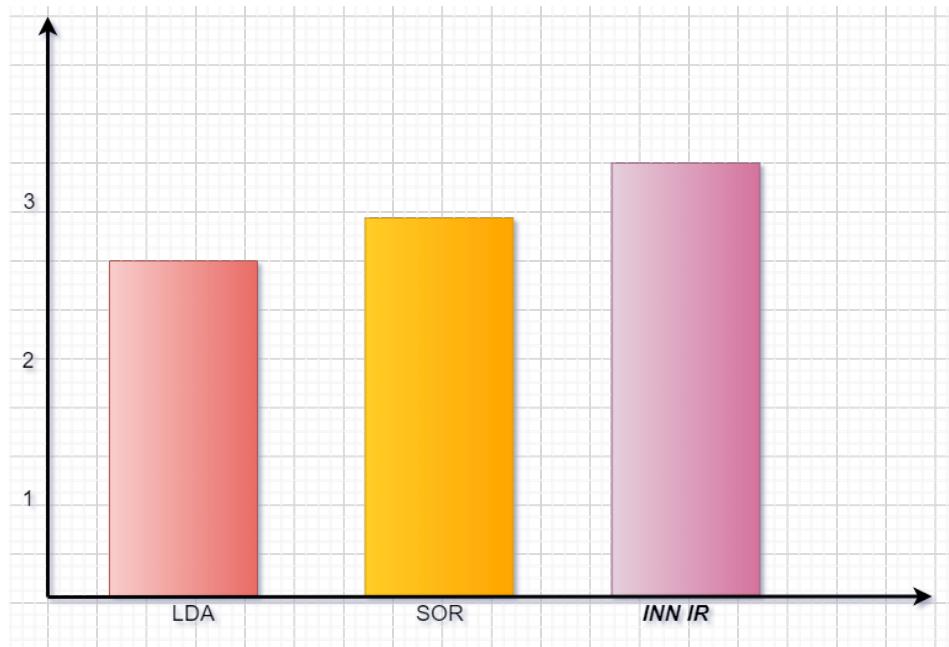


Fig. 5 shows score analysis relevancy

The figure shows that model performs more accurately in different experiments when it is compared with other models. This is because of the use of an intelligent neural network and with the help of page ranking process.

The proposed model utilizes intelligent NN algorithms to categorize web documents into distinct groups based on the relevance of specific words/terms. Additionally, the Neural Network (NN) identifies keywords that are semantically associated with web articles. Experimental findings underscore the importance and relevance of this proposed approach in today's dynamic environment. Furthermore, the comparative study depicted in the graph evaluates various techniques alongside a classifier, demonstrating that our proposed classifier outperforms others, as evidenced by the results displayed.

5 Conclusion and future work

This study proposes the utilization of information retrieval techniques, employing pre-processing and NN-based document classification methods, for semantic information retrieval from user-interest-based social media datasets. Preprocessing technique can remove the useless and raw data from the selected dataset. The Intelligent Neural network technique is used to retrieve texts and compares them to query hypotheses based on user interest. When we compared both syntax-based IR model and semantic-based approaches, these algorithms minimize retrieval time, boost semantic relevancy through the use of NN, and handle large amounts of data more effectively by utilizing an intelligent neural network algorithm that can produce an effective related media tweets for returning more effective similar tweets and retweets of user interest, page ranking is also used for documents ranking which produces better results. Further in future, this approach can include the use of intelligent agents in multiple ways and classifier for the processing of user requests that is very effective.

6 Reference

- [1] Daga Ishita, Anchal Gupta, Raj Vardhana, and Partha Mukherjee. Prediction of Likes and Retweets Using Text Information Retrieval. Procedia Computer Science, 168, pp.123–128, 2020.
- [2] Nesi, Paolo, et al. Assessing the retweets proneness of tweets: predictive models for retweeting. Multimedia Tools and Applications, 77(20), pp.26371-26396.
- [3] P. Mahalakshmi et al. An Art of Review on Conceptual based Information Retrieval. Webology, 18(Special Issue on Information Retrieval and Web Search), January 2021.
- [4] Hahm, G.J., Lee, J.H., & Suh, H.W. Semantic relation based personalized ranking approach for engineering document retrieval. Advanced Engineering Informatics, 29(3), pp.366-379, 2015.
- [5] Jiang, Y. Semantically-enhanced information retrieval using multiple knowledge sources. Cluster Computing, pp.1-2, 2020.
- [6] Furnas, G.W., Deerwester, S., Durnais, S.T., Landauer, T.K., Harshman, R.A., Streeter, L.A., and Lochbaum, K.E. Information retrieval using a singular value decomposition model of latent semantic structure. In ACM SIGIR

Forum, New York, NY, USA: ACM, 51(2), pp.90-105, 2017.

- [7] Furnas, G.W., Deerwester, S., Durnais, S.T., Landauer, T.K., Harshman, R.A., Streeter, L.A., and Lochbaum, K.E. Information retrieval using a singular value decomposition model of latent semantic structure. In ACM SIGIR Forum, New York, NY, USA: ACM, 51(2), pp.90-105, 2017.
- [8] Yu, B. Research on information retrieval model based on ontology. EURASIP Journal on Wireless Communications and Networking, 1, pp.1-8, 2019.
- [9] Nawaz, A., Ali, T., Hafeez, Y., Rehman, S. U., & Rashid, M. R. Mining public opinion: a sentiment-based forecasting for democratic elections of Pakistan. Spatial Information Research, 1-13, 2022.
- [10] Haykin, Simon S, et al. Neural networks and learning machines, Vol. 3. Prentice Hall, USA, 2009.
- [11] Badache, Ismail and Boughanem, Mohand. Document Priors Based On Time-Sensitive Social Signals. In: Lecture Notes in Computer Science, 10.1007/978-3-319-16354-3_68, 2015.
- [12] Surowiecki, James and Silverman, Mark. The Wisdom of Crowds. American Journal of Physics, 75, pp.190-192, 2007.
- [13] Liu, H., Simonyan, K., and Yang, Y. Darts: Differentiable architecture search. arXiv preprint arXiv:1806.09055, 2018.
- [14] Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2(1), pp.1-19, 2006.
- [15] Ahmad, W., and Ali, R. Information retrieval from social networks: A survey. In 2016 3rd International Conference on Recent Advances in Information Technology (RAIT) (pp. 631-635). IEEE, 2016, March.
- [16] Kanhabua, N., Blanco, R., and Nørvåg, K. Temporal information retrieval. Foundations and Trends® in Information Retrieval, 9(2), pp.91-208, 2015.
- [17] Blei, D.M., Ng, A.Y and Jordan, M.I. Latent Dirichlet allocation. J. Mach. Learn. Res., 3, pp.993–1022, 2003.
- [18] Nesi, Paolo, et al. Assessing the retweets proneness of tweets: predictive models for retweeting. Multimedia Tools and Applications, 77(20), pp.26371-26396, 2018.
- [19] Daga Ishita, Anchal Gupta, Raj Vardhana, and Partha Mukherjee. Prediction of Likes and Retweets Using Text Information Retrieval. Procedia Computer Science, 168, pp.123–128, 2020.
- [20] Jingjing Cai, Jianping Li, Wei Li, and Ji Wang. Deep learning model used in text classification, pp.123–126, 2018.
- [21] Lee, D., and Seung, H. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), pp.788–791, 1999.
- [22] Torii, M., Arighi, C.N., Li, G., Wang, Q., Wu, C.H., Vijay-Shanker, K. Rlims-P 2.0: a generalizable rule-based information extraction system for literature mining of protein phosphorylation information. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12(1), pp.17–29, 2015.
- [23] Ledneva, Tatiana et al. Procedia Computer Science, 192, pp.2720–2730, 2021.
- [24] C.S. Saravana Kumar and R. Santhosh. Computers and Electrical Engineering, 81, 106518, 2020.
- [25] Sanz-Cruzado and Castells, 2018a; 2018b; 2019; Sanz-Cruzado, Castells, & López, 2019.
- [26] Li, M., Arighi, C.N., Torii, G., Wang, Q., Wu, C.H., Vijay-Shanker, K. Rlims-P 2.0: a generalizable rule-based information extraction system for literature mining of protein phosphorylation information. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12(1), pp.17–29.
- [27] Shimoda and K. Yanai. Learning food image similarity for food image retrieval. In: IEEE Third International Conference on Multimedia Big Data, 2017.
- [28] Mijangos, Víctor, Sierra, Gerardo, and Montes, Azucena. Sentence level matrix representation for document

spectral clustering. Pattern Recognition Letters, 85, pp.29-34, 2017.

- [29] Premalatha, M., Mathavan, T., Selvasekarapandian, S., Monisha, S., Pandi, D. V., and Selvalakshmi, S. Investigations on proton conducting biopolymer membranes based on tamarind seed polysaccharide incorporated with ammonium thiocyanate. Journal of Non-Crystalline Solids, 453, pp.131-140, 2016.
- [30] Li, M., Arighi, C.N., Torii, G., Wang, Q., Wu, C.H., Vijay-Shanker, K. Rlims-P 2.0: a generalizable rule-based information extraction system for literature mining of protein phosphorylation information. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12(1), pp.17–29.
- [31] C.S. Saravana Kumar and R. Santhosh. Computers and Electrical Engineering, 81, 106518, 2020.
- [32] Sanz-Cruzado, J., Castells, P., Macdonald, C., and Ounis, I. Effective contact recommendation in social networks by adaptation of information retrieval models. Information Processing & Management, 57(5), 102285, 2020.
- [33] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science, 368(6490), pp.489–493, 2020.
- [34] Sakamoto, M., Sasaki, D., Ono, Y., Makino, Y., and Kodama, E. N. Implementation of evacuation measures during natural disasters under conditions of the novel coronavirus (COVID-19) pandemic based on a review of previous responses to complex disasters in Japan. Progress in Disaster Science, 8, 100127, 2020.
- [35] Liu, H., Liu, Y.S., Pauwels, P., Guo, H., and Gu, M. Enhanced explicit semantic analysis for product model retrieval in the construction industry. IEEE Transactions on Industrial Informatics, 13(6), pp.3361-3369, 2017.
- [36] Wei, X., and Croft, W.B. LDA-based document models for ad-hoc retrieval. In: ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06), pp.1–8, 2006.
- [37] Schütze, H., & Silva, M.J. On the role of language modeling in information retrieval. Journal of Computer and System Sciences, 77(6), pp.947–957, 2011.
- [38] Xu, J., & Croft, W.B. Query expansion using local and global document analysis. In: ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96), pp.4–11, 1996.

Article Citation:

Author1_Abraham et al.
(2024). Dual Module
Vibratory Device for
Enhanced Blood Circulation
in Muscles *FUSST*
International Conference on
Computer Science and
Engineering

Dual Module Vibratory Device for Enhanced Blood Circulation in Muscles

Ramis Abraham^a, Hadia Zainab^a, Pakeeza Komal^a, Faisal Ishtiaq^a,
Saad Habib Qureshi^a.

^a Department of Engineering Technology, Foundation University, Islamabad, Rawalpindi, Pakistan.
(ramisabraham60@gmail.com).

^a Department of Engineering Technology, Foundation University, Islamabad, Rawalpindi, Pakistan.
(hadiaa.zainabb@gmail.com)

^a Department of Engineering Technology, Foundation University, Islamabad, Rawalpindi, Pakistan.
(malikpakeeza492@gmail.com).

^a Department of Engineering Technology, Foundation University, Islamabad, Rawalpindi, Pakistan.
(faisal.ishtiaq@fui.edu.pk)

^a Department of Engineering Technology, Foundation University, Islamabad, Rawalpindi, Pakistan.
(saadqureshi.pg@smme.edu.pk)

* **Corresponding author:** Department of Engineering Technology, Foundation University, Islamabad,
Rawalpindi, Pakistan

Faisal Ishtiaq
(faisal.ishtiaq@fui.edu.pk)

Abstract:

The purpose of this project was to design a device that would aid in relaxation by improving circulation in the hand and head. The device utilizes an ESP-32 microcontroller with integrated Wi-Fi to interface with a cloud server, allowing the user to store and access massage data through an email login. The user can customize the runtime and choose to target either the left or right hand using a specialized 6-fingered glove. The hand module employs a lead screw mechanism to mimic the motion of a human hand, targeting 8 key acupuncture points to encourage blood flow. Additionally, a vibrating DC motor attached to the glove's base creates a stimulating pattern. For the head module, a servo motor is coupled with the tightening mechanism of a pre-built helmet, allowing the device to grip and massage the user's head. This innovative device leverages easy-to-understand mechanisms to provide a comprehensive relaxation experience, with the added benefit of cloud-based data storage and access.

Keywords: Relaxation; Circulation; 6-fingered glove.

1. Introduction

In recent years, the development of innovative therapeutic devices has become increasingly important for enhancing patient outcomes, particularly in the areas of circulation improvement and disability management. This paper introduces a novel dual-module vibratory device specifically designed to address these challenges. The device combines two distinct modules: one aimed at enhancing blood circulation through targeted vibration therapy, and the other focused on alleviating disability-related symptoms by providing customized vibrational interventions. Circulatory blood flow in the body is either achieved by the vascular pressure inside the blood vessels created by the heart or the movement of the skeletal muscles allowing the smooth flow. External factors such as age and lifestyle hinder the optimal self-sustained process set by the human body. Vibration therapy and external aid from our device allows the human body to adverse the limited affects and relief the user from fatigue and suspended blood flow. The device features physical forcible apparatus mimicking the mechanism of that of another human, certain gestures that over the years have been therapeutic and acted as a home remedial methods of treating such inconveniences. By integrating these functionalities into a single platform, this device offers a comprehensive solution that could significantly improve the quality of life for individuals with circulatory issues and disabilities. External forcing and contact with the body forces the muscles to act normally pushing blood encouraging movement throughout the body, reducing physical stress from the body helping in tackling long termed set diseases. The following sections explore in detail the design, implementation, and clinical evaluation of this dual-module device, demonstrating its potential as a transformative tool in therapeutic care.

Copyright
Copyright © 2024
Author1_Abraham et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

potential as a transformative tool in therapeutic care.

Evolution of the research field:

There are over 50 muscles in the hand which work in coordination, this repetitive motion to the muscles and tendons of the hand causes strain and leads to injuries. Reduced blood flow caused by poor circulation and diseases and health deterioration of the human body, numerous conditions such as ischemia, peripheral vascular disease and impaired wound healing. Double crush syndrome [1] which affects the wrist and the spinal nerves in the neck.

Abraham et al. "Dual Module Vibratory Device for Enhanced Blood Circulation in Muscles"

The vertebral column or the spinal cord is made of fibro cartilaginous [2] intervertebral discs. These structures are covered with nerves and help in movement and maintains the posture of the body. It is crucial to recognize the role of proper blood circulation in maintaining hand health. Therefore, interventions to enhance circulation are of paramount importance for preventing and managing these issues. The more common type of issue faced is nerve compression [3] due to long stagnant movement by the human muscles causing stiffness and extreme pain afterwards due to low blood flow rendering the muscle completely ail for use.

The research gap and its importance:

While vibration therapy has shown promise in improving circulation in isolated body regions, there is a lack of specialized devices and corresponding research that target the simultaneous treatment of the hand and neck regions. The combined application of vibration therapy to these areas has not been adequately explored, particularly in terms of tailored, customizable approaches that account for the unique anatomical and physiological needs of these regions. Moreover, there is a scarcity of long-term studies evaluating the safety and efficacy of such dual-module devices, highlighting the need for further research to fill these critical gaps. Long-term studies on the safety and effectiveness of vibration therapy in these specific regions, particularly when using a dual-module approach, might be sparse or incomplete.

Brief description of the proposed solution:

It is crucial to recognize the role of proper blood circulation in maintaining hand health. Therefore, interventions to enhance circulation are of paramount importance for preventing and managing these issues. Circulation devices, focusing on their design, functionality, and clinical applications provide relief by improving blood circulation, manage vascular conditions, and enhance overall hand health. Massage leads to improved body circulation and relaxation to the muscles, along with specific vibration being applied to the areas. Vibration therapy [4] has proven to be promising by targeting the specific regions resulting in the proposed outcome to be positive.

Key Contributions of the Study:

- The use of a lead screw mechanism to mimic the motion of a human hand and target specific acupuncture points is a unique approach that could be further explored and refined.
- Integrating the vibrating DC motor to enhance blood flow stimulation is an innovative addition to the hand massage functionality.
- The integration of a servo motor with the helmet's tightening mechanism to provide an adjustable and customizable head massage is a novel feature.
- Exploring the optimal grip pressure and massage patterns for the head module could lead to improvements in user comfort and efficacy.
- Enabling cloud-based data storage and access through email login is an interesting way to enhance the user experience and facilitate long-term tracking of massage sessions.
- Investigating the potential benefits of cloud integration, such as remote monitoring, data analysis, and personalized recommendations, could be valuable research directions.
- Evaluating the user experience, comfort, and ease of use of the 6-fingered glove and helmet design could lead to insights for improving the overall ergonomics of the device.
- Conducting user studies to understand the device's effectiveness in promoting relaxation and improving circulation could validate the targeted design approach.
- Exploring the potential applications of this device in fields such as physical therapy, sports recovery, or stress management could uncover new research avenues and expand the device's impact.
- Development of a Dual-Module Device: Introduced a novel device combining two modules specifically designed to improve blood circulation and relieve disabilities through targeted vibration therapy.
- Innovative Vibration Mechanism: Designed and implemented an innovative vibration mechanism that adapts to individual patient needs, offering customizable therapeutic options.
- Clinical Evaluation: Provided the first comprehensive clinical evaluation of the device's efficacy in improving circulation and reducing disability-related symptoms, demonstrating significant benefits over existing solutions.
- Design Framework: Established a new design framework for medical devices aimed at multi-functional therapeutic interventions, which can be applied to future research and development
- Collaborating with experts in areas like acupuncture, massage therapy, and biomedical engineering could lead to further refinements and innovations.

1.1. Research Objectives

The primary research objective of our dual-module vibratory device is to develop and evaluate a comprehensive therapeutic solution that effectively addresses chronic forearm muscle stiffness and tension headaches resulting from prolonged and stressful work conditions. Given the increasing prevalence of these conditions in today's fast-paced and often ergonomically challenging work environments, there is a pressing need for an innovative device that not only alleviates discomfort but also enhances overall well-being and productivity.

Abraham et al. "Dual Module Vibratory Device for Enhanced Blood Circulation in Muscles"

Traditional therapeutic options, while beneficial, often fall short in delivering a targeted and integrated approach to treating both the hand and neck regions simultaneously. Many existing products are either limited to treating isolated areas or lack the customization required to address individual needs comprehensively. Furthermore, the practicality of current devices can be problematic, as they may be cumbersome or fail to provide adequate real-time feedback on the user's condition.

1.1.1. Novel concepts centered for improvement

Medical and technological advancements have allowed to refine the already devised methodologies simple yet effective ideas pinpointing minute details of the human body by boosting the need for in depth technicalities.

Acupuncture [5] a traditional Chinese medicine practice that involves inserting thin needles into specific points on the body to stimulate energy flow and promote natural healing. Rooted in ancient Chinese philosophy, apart from the physiological beliefs there is proof of acupuncture [6] providing stimulation by influencing the nervous system by stimulating nerve fibers, which then trigger the release of neurotransmitters such as endorphins and serotonin. These chemicals help modulate pain and enhance mood, providing relief from chronic pain conditions such as arthritis, back pain, and migraines. Further reduces inflammation by affecting inflammatory cytokines and other mediators. This can be beneficial for conditions characterized by chronic inflammation, such as osteoarthritis or autoimmune diseases. Needle insertion stimulates local blood flow, which can aid in tissue repair and the delivery of oxygen and nutrients to injured or stressed areas. Improved circulation supports the body's natural healing processes and can expedite recovery from injuries.

Our device targets the two set main aspects of the human body consisting of the most number of pressure points [7] that are easily accessible and can be applied to treatment without hindering their lifestyle or going through a rigorous setup or change of environment. We chose the palm of the hand and the neck region targeting the pressure points of the palm. The head module consisted of vibrational techniques allowing the second to most drained muscle region when a person undergoes strict and stress full working conditions that limits bodily movements.

1.1.2. Market landscape and differentiation.

Devices available in the market target a single feature and forcing the user to buy specific devices for each of the problem they are facing. Additionally, many of these products are either too cumbersome for regular use or do not offer the level of customization needed to cater to individual preferences and varying levels of discomfort. In contrast, the development of a dual-module vibratory device that targets both the hand and neck regions stands out by providing a comprehensive and user-friendly solution. Keeping in mind the need of including two major regions that would highly induce the effect of this therapy led us to two device modules, offering the user with optional and vital coverage of the assistance that we aim to provide.

Furthermore, the device's ability to customize vibration intensity and pattern provides a tailored therapeutic experience, accommodating the diverse needs of users with varying levels of discomfort. The inclusion of a lead screw mechanism simulates the effects of manual massage, offering a more natural and effective treatment option compared to standard vibration alone. Many devices are offered as a consumer product and being irrelevant towards the demand of the consumer our product sets out different from those by offering an engaging experience. Additionally, the device is designed with user convenience in mind, incorporating a screen display for real-time user feedback and monitoring, a sliding drawer lever for easy hand insertion, and a pulse oximeter for heartbeat checks before and after the massage. The device gives the user the choice of operating both modules simultaneously or at a time according to the relative need of the user. These features not only enhance the therapeutic benefits of the device but also make it more practical and user-friendly for everyday use, setting it apart in a crowded and competitive market.

By addressing the specific challenges that existing products fail to meet, this dual-module vibratory device offers a differentiated and superior solution, ultimately improving the quality of life and productivity for individuals suffering from chronic work-related muscle stiffness and tension headaches.

1.2. Research Contributions

The development of the dual-module vibratory device represents a significant advancement in the field of therapeutic devices aimed at improving circulation and alleviating disabilities. Many have explored in the field of vibration therapy such as the effect it has on blood flow [6]. This article helped us understand and narrowing the aspects of our device we needed to engineer. The main aspects of including the whole of vibration therapy has positive effects and if any. This research paper also led us to develop the head module as improved blood flow in the head was vital.

Conceptual and Empirical contributions:

Our research led us to find existing ideas to be proving effective for our intended purpose. Such as lead screws which allowed us to target the pressure points of the palm, however a fix mechanism was not available. This led to the making of the lead screw mechanism.

The lead screw mechanism is a basic concept that we refined furthermore according to our needs.



Figure 1: Coupling of the lead screw and the shaft motor



Figure 2: Complete Lead Screw Mechanism.

This worked together with a stepper motor, a mechanical shaft and the lead screw. We housed this mechanism in a small pipe for better fitment. This helped in securing them to the body as they needed to be robust and precise while performing the therapy to the hand. They undergo the most movement and come in contact and therefore required to be majorly important for the purpose of the device.



Figure 3: PVC holdings for our LSM's

Abraham et al. "Dual Module Vibratory Device for Enhanced Blood Circulation in Muscles"

While designing innovative features that further improved simply the operating of the device were largely considered and were extremely well thought. A sliding drawer which allowed comfortable placement of the hand of the user, minute details led us to the better design of the device and providing a better experience to the user. A head band for the helmet module gripping the head and securing the device as the sudden change in mood and can allow the user to change his or her orientation resulting in an effective use of the device.

A pre marketed helmet which housed the mechanism for the head module, this included the electrical and mechanical components.



Figure 4: Head module

Another breakthrough design was the 6-fingered glove which allowed either hand to be placed certainly in the device for the therapy, having the thumb on either side of the single glove, this was useful in eliminating the need for changing the glove every time for hand change.



Figure 5: 6-fingered glove

2. The Research Problem

Chronic forearm muscle stiffness and tension headaches are increasingly common issues among individuals subjected to prolonged stressful work environments, particularly those who spend extended periods in repetitive tasks or static postures. These conditions not only cause significant physical discomfort but also lead to a notable decline in quality of life, as they interfere with daily activities and reduce overall productivity. The repetitive strain on the forearm muscles often results in stiffness, making even simple movements painful and difficult, while tension headaches, frequently triggered by stress and poor posture, add to the discomfort and cognitive fatigue. Despite the availability of various therapeutic options, many existing solutions are either inadequate in addressing the specific needs of these conditions or are too inconvenient for regular use, particularly in a fast-paced work setting. What is urgently needed is a more effective and user-friendly therapeutic solution that can simultaneously target both the hand and neck regions—two critical areas commonly affected by work-related stress. Such a solution would not only alleviate the physical symptoms but also enhance overall well-being and productivity by providing a holistic approach to managing the cumulative effects of chronic stress on the body. This underscores the importance of developing innovative devices that can offer tailored, convenient, and comprehensive treatment, ultimately improving the quality of life for individuals dealing with these debilitating conditions.

3. Proposed Solution

A dual-module vibratory device is designed to alleviate forearm stiffness and tension headaches by combining a 12V DC motor with an unbalanced weight to generate customizable vibration patterns and a lead screw mechanism to

Abraham et al. "Dual Module Vibratory Device for Enhanced Blood Circulation in Muscles"

simulate manual hand massage. This was upheld by the vibration therapy proving to be well effective with a few tweaks. Designing the device to be better at what the research had concluded in order for that we came up with original ideas that were self-inspired by team members, allowing to push the limitations of existing products. Housing all the ideas under a single settlement, devising the product into two separate modules allowing the user to freely choose between the said therapy of the hand or the neck and head.

Innovations allowed us to be exact at the targeted purpose, we redesigned the lead screw mechanism that targeted the pressure points of the palm of the user which are the main basis of our therapy research. These involved the use of computer language and electric industrial components such as limit switches to help us in achieving the desired result. A 6-fingered glove design allowing the user to house both hands for smooth and a comfortable experience.

The device allows users to select the hand and adjust vibration intensity, provides real-time information on a screen display, and monitors heart rate with a pulse oximeter. This shows the effectiveness of the device after proper usage. Other innovative ideas include a sliding drawer lever for easy hand insertion. A vibration net for the head module targeting the whole head and the cervical region which is the neck

Furthermore, our device houses two separate modules, this leaves room for improvement on the designing phase next. Making it compact and centered around portability which is major demand nowadays reducing the need for heavy machinery and load.

4. Results and Discussion

The purposed device indulged idea behind the descriptive research was displayed in a manner of public testing manner. The user toggled and navigated through the device accumulating the experience to that similar to that of operating a large hospital device rather than a palm sized massager. This allowed the user to be in control and fascinated by the device further indulging the users mind to interact with the device. The major turning point was the use of the device and the ability to track the need of the user, thus staying true to the expectations they has from the device. While designing the interface we had certain self-indulging human behavioral pattern studied which allowed the user to more curious and finishing the setup of the device for its desired use. At every step of the way it made the user to be more involved and connected with the device.

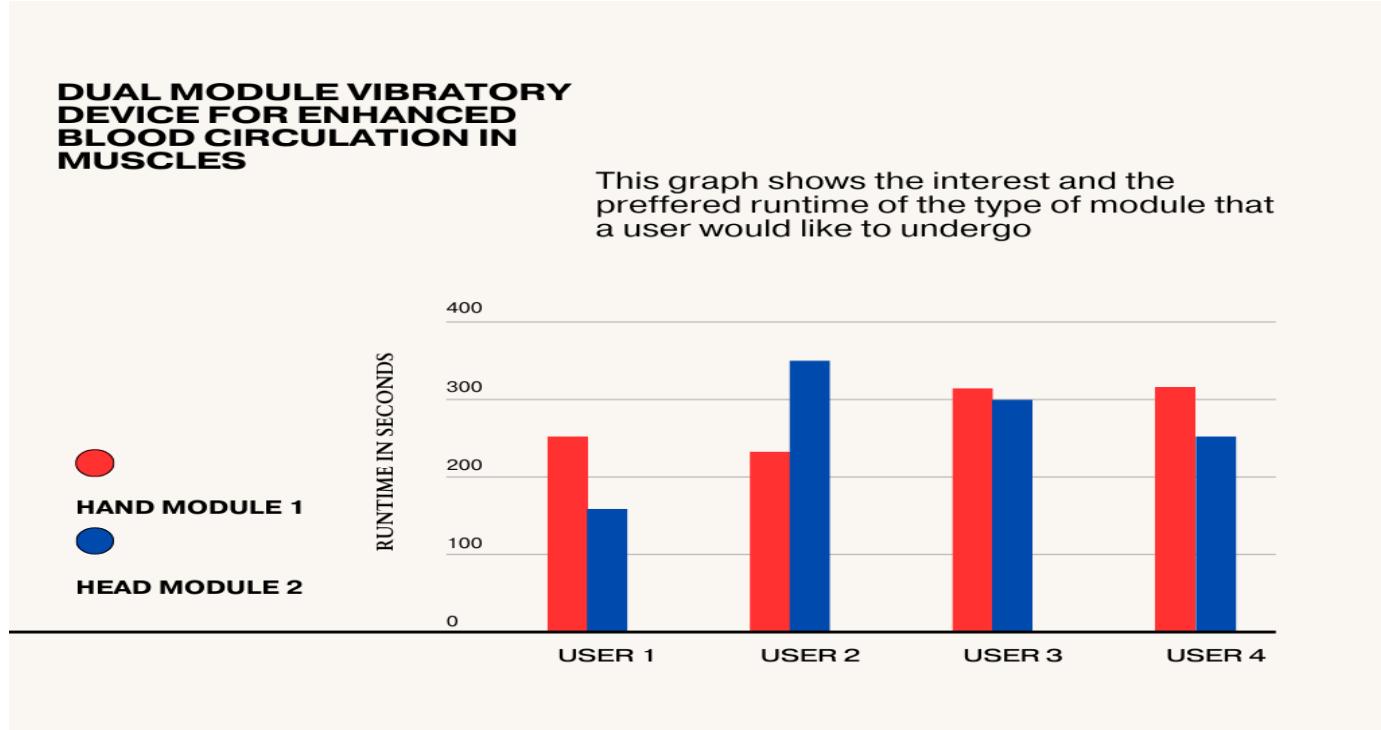


Figure 6: Testing phase and feedback analysis

5. Conclusions and Future Direction

The development and implementation of our assistive circulation device for hand and the head hold tremendous potential in enhancing the well-being of individuals facing circulatory challenges. The assistive circulation device serves as a promising solution for individuals dealing with various circulatory issues in the hands, providing targeted relief for conditions such as arthritis, swelling, or reduced blood flow.

Our device is setup for undergoing scrutinized design changes, reducing the extra space and making it minimalist and proving to be more ergonomic than before, we discussed with a software centered group which will help us in making the interface more user friendly and be centered on the needs of the user.

Acknowledgment

Furthermore the input from the research reviewers guiding throughout the writing and research of the topics are extremely recognized as their input played an essential role.

Conflict of Interests

This research article has no conflict of interest from any publication.

Funding Statement

This well researched document is under no obligation from any type of external funding.

6. References

- [1] J. N. Osterman, L. R. Seibert, A. D. Cronin, R. T. Calkins, and M. M. Stern, "Double Crush Syndrome," *J. Am. Acad. Orthop. Surg.*, vol. 23, no. 9, pp. 558-562, Sept. 2015.
- [2] J. B. Benjamin and J. R. Ralphs, "Fibrocartilage in tendons and ligaments: An adaptation to compressive load," *J. Anat.*, vol. 193, no. 4, pp. 481-494, 1998.
- [3] A. A. Pearl and J. R. Eaton, "Pathophysiology of vibratory hand tools," *Hand Clinics*, vol. 17, no. 2, pp. 235-240, May 2001. [Online]. Available: [https://www.hand.theclinics.com/article/S0749-0712\(01\)00012-9/abstract](https://www.hand.theclinics.com/article/S0749-0712(01)00012-9/abstract). [Accessed: Aug. 28, 2024].
- [4] C. E. Clarkson, "Hand-arm vibration syndrome: A historical perspective," *Occupational and Environmental Medicine*, vol. 49, no. 1, pp. 1-2, Jan. 1992. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1743552/>. [Accessed: Aug. 28, 2024].
- [5] M. J. Fichter and R. J. Frush, "Reduction of radiation dose for pediatric body MDCT: time to think outside the box," *Mayo Clinic Proceedings*, vol. 88, no. 2, pp. 127-130, Feb. 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0025619613005132>. [Accessed: Aug. 28, 2024].
- [6] M. M. Figueiredo, P. A. V. Farias Neto, L. T. de Oliveira, and J. P. Sousa, "Evaluation of vibration therapy in reducing pain and improving functionality in patients with chronic low back pain: a randomized clinical trial," *J. Bodyw. Mov. Ther.*, vol. 24, no. 1, pp. 66-72, 2020.
- [7] L. A. A. MacPherson, J. A. Lewith, G. Richard Hammerschlag, and D. C. Kaptchuk, "Point specificity in acupuncture," *Chinese Medicine*, vol. 7, no. 4, p. 4, 2012. [Online]. Available: <https://doi.org/10.1186/1749-8546-7-4>



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

IoT-Driven Hand Gesture Recognition for Intelligent Home Automation Using Landmark Detection

Baleegha Fatima^a, Bilal Mushtaq^{a,*}, Abbas Ahmed^a, Ammar Iqbal^a

^a Department of Electrical and Electronics Engineering, Beaconhouse International College, , Islamabad, Pakistan.

(baleegha.8197@bic.edu.pk).

(Muhammad.ammar.98330@bic.edu.pk)

(bilal.mushtaq@bic.edu.pk & Bilalmushtaq88@outlook.com).

(abbas.ahmed@bic.edu.pk).

* Corresponding author: bilalmushtaq88@outlook.com

Abstract:

Article Citation:
Fatima et al. (2024). "Hand Gesture Based Home Automation Using IoT". FUSST International Conference on Computer Science and Engineering

This paper offers a comprehensive examination of advancements in assistive technology tailored for deaf individuals with hard hearing impairments, with a specific focus on the role of the Internet of Things (IoT) in enhancing their quality of life. Sign language technologies are important in the life of Individuals with Auditory Challenges, but more issues remain to be tackled like likelihood of communication with non-signing people and no automated recognition tools are available. This paper describes implementation of a video based American Sign Language identification system that employs MediaPipe for hand tracking, OpenCV for image normalization and Gesture Control based Convolutional Neural Network (CNN) for localization of the gestures. Written in Python, the application records video streams, filters out the hand regions and recognizes ASL letter gestures. Operating at high recognition rate and accuracy in home automation, the system utilizes computer vision and machine learning approaches to increase recognition and improve users' experience with the system. This research breaks communication barriers and enhances the inclusivity of people with physical and sensory disabilities by supporting the access and additional technologies.

Keywords: Convolutional Neural Networks (CNN), Sign Language, Artificial Intelligence (AI), Machine Learning (ML), American Sign Language (ASL), Image Processing.

1. Introduction

Gestures, including movements of the hands, fingers, and arms, are fundamental for conveying information in human interactions [1][2][3][4]. For individuals who rely on sign language, gesture-based inputs provide an empowering method for controlling devices, significantly enhancing accessibility and autonomy [5].

In the field of human-computer interaction, two primary approaches dominate: "data gloves" and "vision-based" methods. This study adopts the vision-based approach, focusing on hand gesture recognition and classification integrated with an Arduino-UNO microcontroller for home automation. Hand gestures are crucial in enhancing user interaction with devices by tracking hand movements, which involves computer vision tasks such as hand segmentation, part detection, and tracking. 2D cameras are particularly useful in capturing finger movements, which occur on a 2D plane [6][7][8].

The study, titled "Hand Gesture Recognition of ASL Model-Based Home Automation in Python," addresses the need for inclusivity by integrating advanced home automation technologies with computer vision. Gesture-based interaction can significantly impact home automation, particularly for individuals with physical disabilities or movement restrictions where traditional controls may be difficult to use. The CNN-based method



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024
Baleegha Fatima et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

employed in this study is believed to enhance the accuracy of gesture recognition [9]. However, challenges such as varying illumination levels and other environmental factors still affect recognition performance [10]. Additionally, there is an urgent need for more ethnically and culturally representative data to ensure accurate gesture identification across diverse user groups [11]. Despite these challenges, hand gesture recognition technology holds great promise for transforming how people interact with devices.

The history of the modern smart home concept of an integrated household appliance as a room in the vestiges of a 20th-century society brings out the trend of technology in the domestic sphere [15]. Noteworthy advancements like ECHO IV and Kitchen Computer can be seen as the first steps in smart home technology progression [15][16] [17]. The 2000s witnessed the advent of affordable cost smart homes that assured optimal utility and privacy through advanced security mechanisms operated at specified frequency [18][19] especially to people with disabilities or short limbs [20]. Figure 1 illustrating the key developments in communication, gesture-based inputs, human-computer interaction systems, and smart home technology.

Early Human History: Two aspects of human life were established

- **Communication** recognized as a vital aspect of human life [1].
- **Gestures** such as hand, finger, and arm movements become essential for conveying information [2].

19th Century: In this era two main aspects were developed

- **Development of Sign Language:** Emergence of sign language as a crucial communication method for individuals with hearing impairments[3].
- **Exploration of Gesture-Based Inputs:** Early exploration into using gestures for enhancing accessibility and autonomy [4].

Late 20th Century: In this time two main things were established

- **Introduction of Data Gloves:** Devices enabling gesture tracking to improve human-computer interaction [5].
- **Vision-Based Methods:** Focus on hand gesture recognition and classification using computer vision techniques [6].

1960s-1970s: In this decade following ways were adopted for the improvement of life

- **ECHO IV (1966):** One of the earliest home automation systems, marking the beginning of smart home technology [7].
- **Kitchen Computer (1969):** An early attempt at integrating technology into domestic life [7].

2000s: In these era technology had emerged for the special people

- **Affordable Smart Homes:** Emergence of homes offering advanced security and utility features, particularly benefiting those with disabilities [8].

2010s: In this era machine learning approaches were developed for human reorganization

- **CNN-Based Gesture Recognition:** Introduction of CNN-based methods for improved accuracy in gesture recognition, with a focus on home automation [9].

2020s: machine learning algorithms were integrated with IoT devices to enhance the development of final products.

- **Key contributions during this period include Real-Time Hand Gesture Recognition Systems:** Integration of gesture recognition with Arduino-UNO and home automation systems, aimed at enhancing accessibility for individuals with hearing impairments [10]-[18].

Figure 1: Timeline of Key Developments in Communication, Gesture-Based Interaction, and Smart Home Technology.

This paper explores the latest advancements in hand gesture recognition systems for home automation, with a particular focus on integrating gesture recognition with home automation technologies. By employing a hand landmarks model to track key points on the hand, the system can infer the gestures being performed. Our approach to home automation through real-time hand gesture recognition exemplifies advanced machine learning and human-computer interaction, enabling users to intuitively control devices and systems. This system has the potential to revolutionize domestic technology interactions, making them more accessible and user-friendly for applications such as home automation, accessibility enhancement, elderly care, and emergency response.

Furthermore, we suggest that future advancements in technology will facilitate further enhancements in user experience and the development of new applications. This project aims to create a real-time hand gesture recognition system for ASL and interface it with home automation systems. The ultimate goal is to design a simple yet effective communication interface for individuals with hearing loss, which would also improve the coordination and efficiency of home automation systems [12][13][14].

1.1. Research Objectives

The primary aim of this paper is to develop a system that can recognize hand gestures in sign language in real time and implement this system in hardware to facilitate communication. The work is evaluated for its alignment with the Sustainable Development Goals (SDGs) [24], highlighting its strengths and identifying areas for improvement. This

research notably contributes to the following SDGs:

- Industry, Innovation, and Infrastructure
- Sustainable Cities and Communities

The significant contributions of this paper include:

- Utilization of Landmark Detection: The system leverages machine learning techniques, specifically using "MediaPipe" and "OpenCV," implemented in Python, to detect landmarks on the hand.
- Real-Time ASL Gesture Recognition: The algorithms are trained to accurately interpret American Sign Language (ASL) gestures in real time using machine learning methods.
- Hardware Implementation: The project establishes and develops a hardware interface for real-life applications aimed at assisting individuals with special needs, utilizing an Arduino-UNO microcontroller.

2. Methodology

Gestures play an essential role in our lives, and without them, life would be challenging. To address this, we have developed a system that recognizes hand gestures using OpenCV, Python, and MediaPipe. The system captures images of the user's hand gestures through a camera, which are then analyzed using the MediaPipe framework in real time. The processed gestures are used to control various home appliances, such as fans and lights, through an Arduino-UNO module. The entire system is managed and controlled by the Arduino-UNO microcontroller.

While vision-based systems are commonly used for hand tracking, they encounter significant limitations, such as restricted movement areas and challenges when hands are crossed or partially visible. Gesture recognition systems, as depicted in Figure 2, can detect real-time movements in multiple directions. By segmenting the hand using color and depth data, the system isolates key parts, such as the palm, wrist, and fingers, while excluding the arm and wrist since they contribute minimal relevant information. The system then analyzes essential features like finger position, height, and distance from the hand's center, ultimately extracting and recognizing the critical components of the gesture.

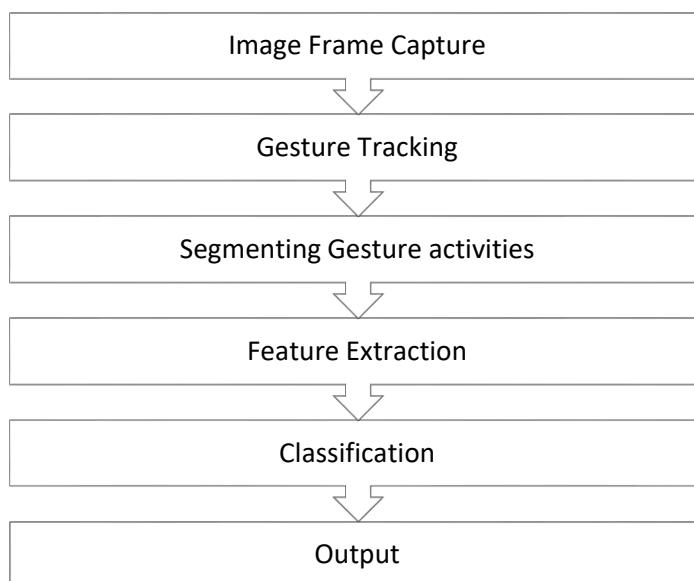


Figure 2: Frame Capture to Output Sequence

From Figure 2 it can be deduced that the process begins with frame capture, where images are obtained from a video feed. These frames provide the raw data necessary for subsequent analysis. Tracking involves monitoring the position and movement of objects within the frames to make necessary assessments. Once tracking is established, the system can isolate gestures from the images and various objects. Feature extraction identifies relevant properties, such as gesture shape and motion, and determines their context. Classification then organizes these features into compact forms, following set rules to define and categorize the gestures. At the output stage, results are provided as graphical displays, text, or commands corresponding to the recognized gestures.

The hand gesture identification method starts with locating the hand in the image, followed by hand refinement, feature extraction, and feature mapping to a gesture dictionary. The goal is achieved by processing the captured gesture. After

the features are extracted, 'pyfirmata' is used to control an Arduino UNO board by relaying the identified features into action, as depicted in Figure 3.

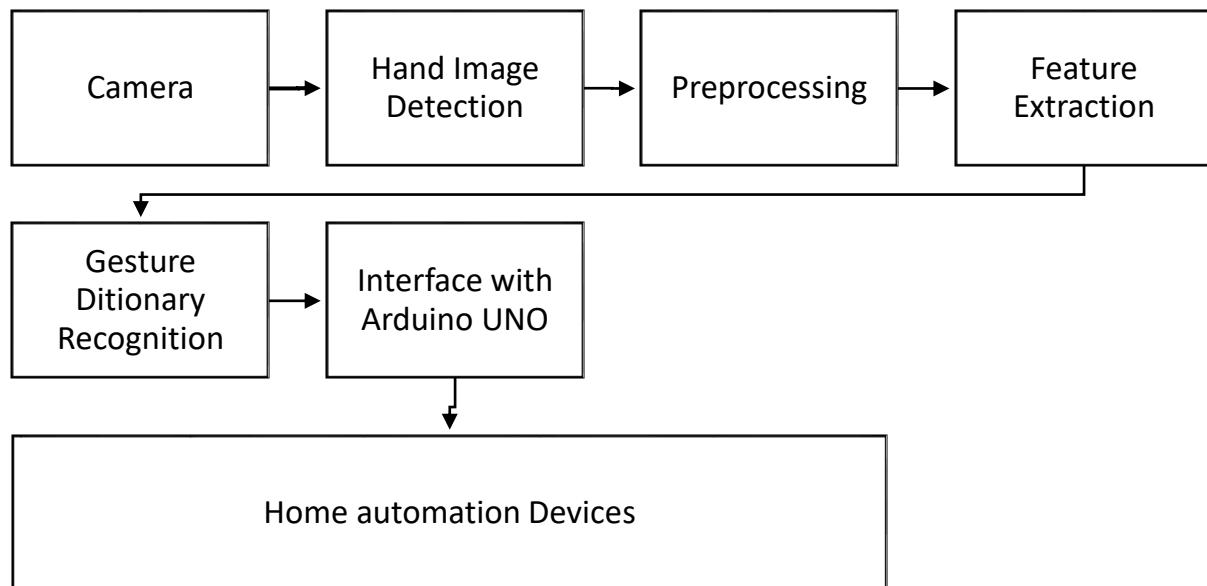


Figure 3: Gesture Recognition and Automation Execution

The Hand Landmarks model is capable of estimating and outputting 21 coordinate points corresponding to knuckles and key regions of the hand. These points include finger joints, knuckles, the palm center, and the wrist, which assist in tracking and analyzing minor motions and hand positions. These coordinates provide accurate location data, forming the foundation for image and gesture recognition, as illustrated in Figure 4.

In video or live streaming mode, the Hand Landmarker leverages the tracked palms from the previous frame, using the bounding box generated by the landmarks model to identify hands in subsequent frames. This approach reduces the workload for palm detection. Palm detection only reactivates when the landmarks model fails to detect or recognize a hand.

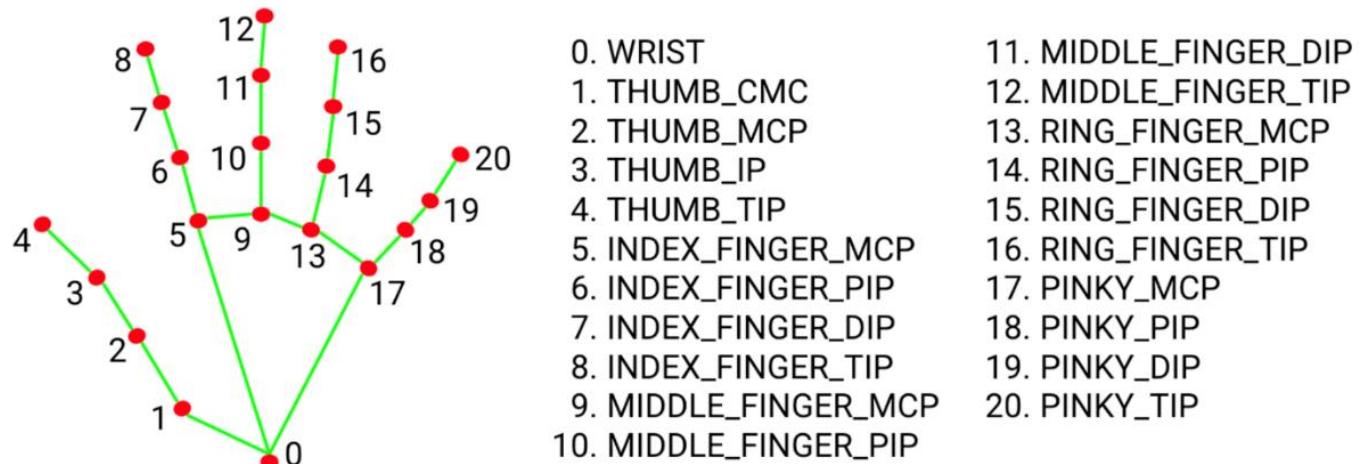


Figure 4: Hand Landmark Model

To achieve real-time video compression, Mediapipe is used due to its ability to perform multiple processing tasks concurrently. It utilizes a bounding box for hand tracking, which efficiently reduces the frequency with which the palm detection model is employed. The system captures sequences of RGB images, accelerating the process by focusing on the temporal dynamics of hand movements.

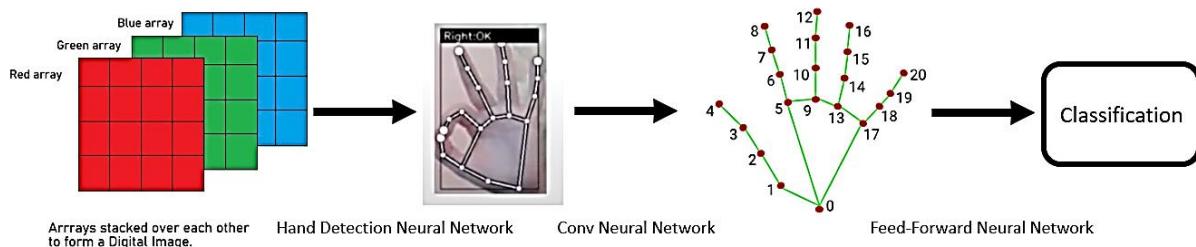


Figure 5: Neural Network Classification Sequence

A Hand Detection Neural Network (HDNN) is implemented to isolate the hand from the rest of the body. A Convolutional Neural Network (CNN) then applies the landmarks to regions such as the palm curve, knuckles, and fingertips to capture spatial information about the hand's shape. This information is further processed and classified using a Feed Forward Neural Network (FFNN), enabling the recognition of complex gestures as the network's architecture is elaborated.

The accurate recognition of hand gesture poses is still a difficult task due to several aspects like small size of dataset. In this paper we selected a large database of size 1000000 accuracy and effectiveness of CNN framework. According to the number of alphabets, we have divided alphabets into 26 classes from 0 to 25. We use webcam of laptop to capture the hand moments.

Table 1: Data Set

Data Set	Total images	Batches/classes	Images in a batch/class
Data Set	1000000	26	38461

Training Dataset = 80000, 80% of Dataset

Test Dataset = 20000, 20% of Dataset

Total params = 1,669,290

Loss Formula $L = -\sum_i^n p_i * \log(q_i)$ where n= No. of classes (26) and p_i is true probability of ith class in the target and q_i is softmax probability of ith class.

$$\text{Accuracy} = \frac{\text{TRUE}_P + \text{TRUE}_N}{\text{TRUE}_P + \text{TRUE}_N + \text{FALSE}_N + \text{FALSE}_P}$$

Table_ 2: Character batches and Accuracy

Characters	Accuracy (%)	Characters	Accuracy (%)
A	88	N	83
B	89	O	84
C	87	P	85
D	84	Q	86
E	82	R	88
F	84	S	80
G	83	T	80
H	85	U	87
I	83	V	83
J	80	W	82
K	85	X	80
L	88	Y	80
M	83	Z	83

3. Results:

The engagement of specific gestures, such as 'A', 'B', and 'C', enables real-time control of system actions, as illustrated in Figure 6. For example, the ASL sign language gestures for 'A', 'B', and 'C' can be translated into actions like operating

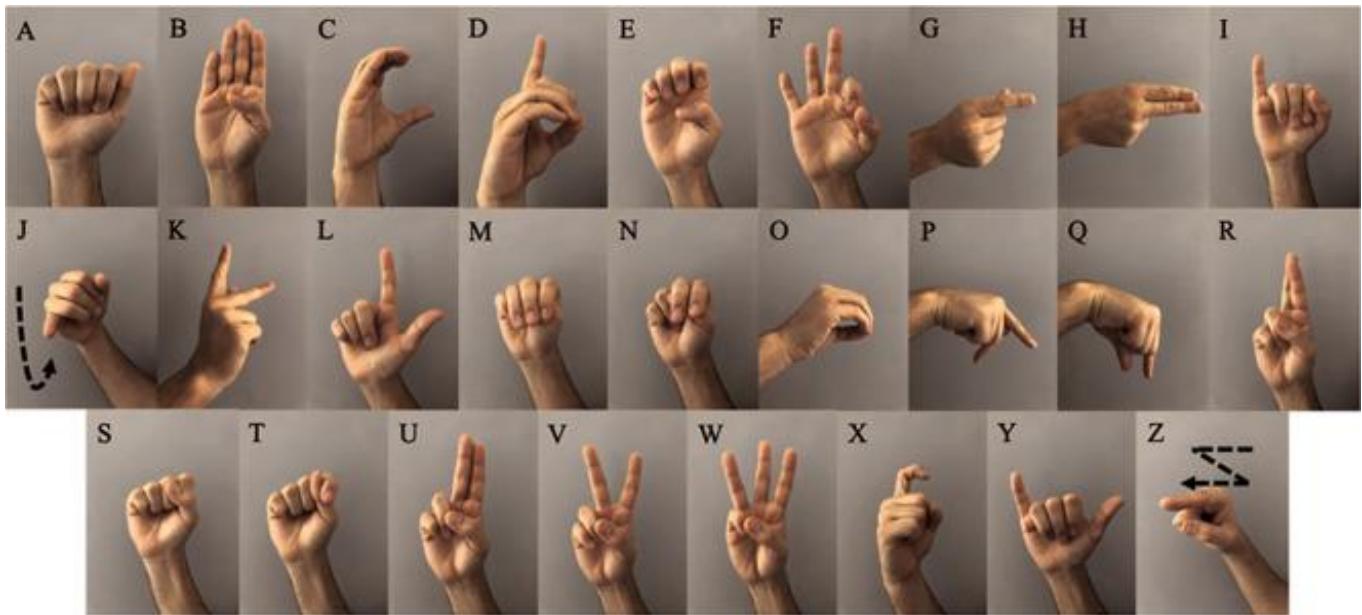
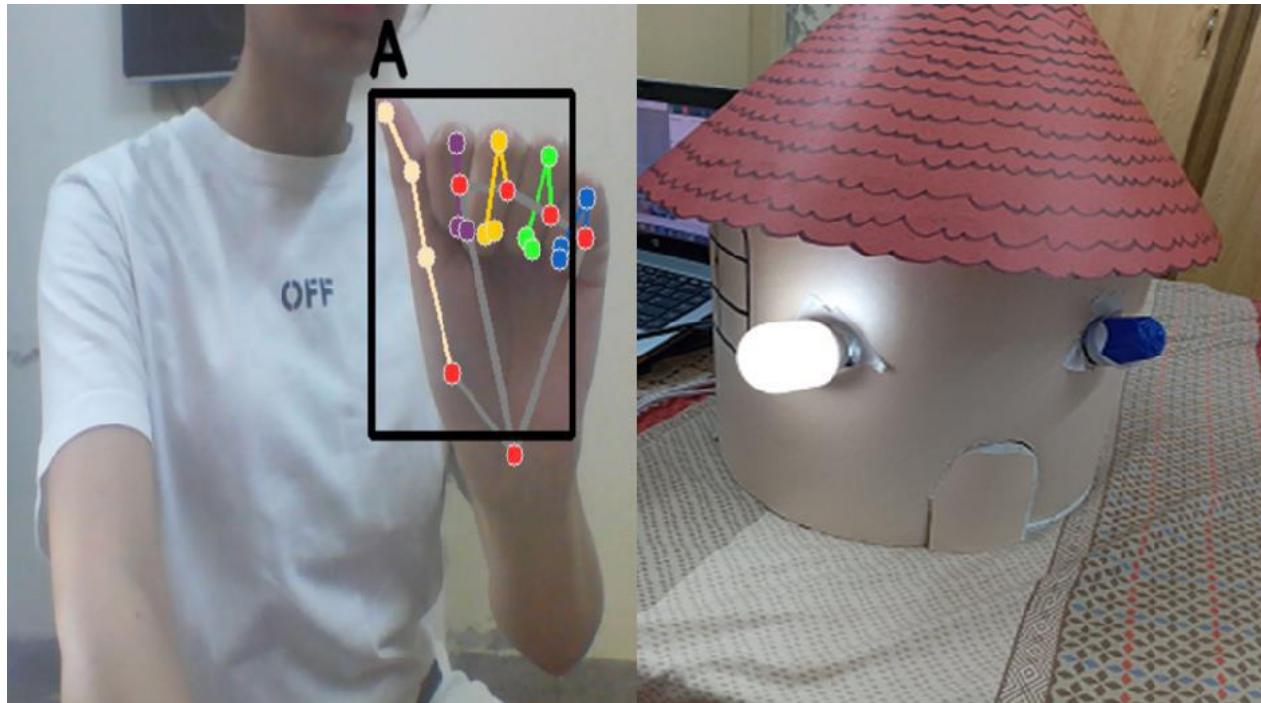


Figure 6: American Sign Language (ASL) Gestures



a light bulb or activating a servo motor for a door mechanism. The hand signal 'A', shown in Figure 7, triggers the operation of a standard white light. This gesture is detected by the camera, and the light turns on instantly.

Figure 7: ASL Gesture "A" and white bulb

Figure 8 illustrates the process of identifying, recognizing, and subsequently switching on the lights. To activate two lights, a hand motion labeled 'B' was required. Both actions are facilitated by a camera in the system, which is capable of detecting both white and blue lights, as depicted in Figure 8. This demonstrates the system's capability to perform multi-stage tasks using the corresponding ASL hand motions.

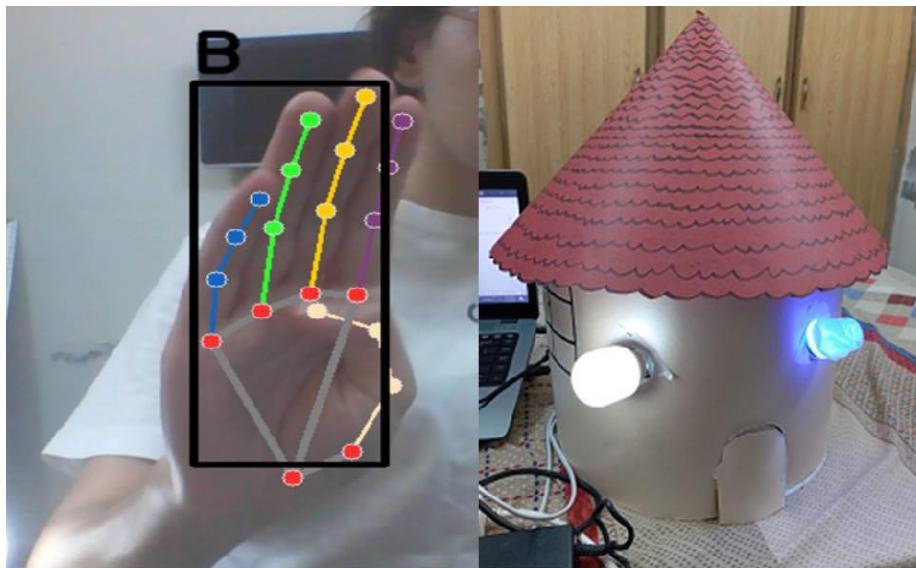


Figure 8: ASL Gesture "B" both white and blue bulbs

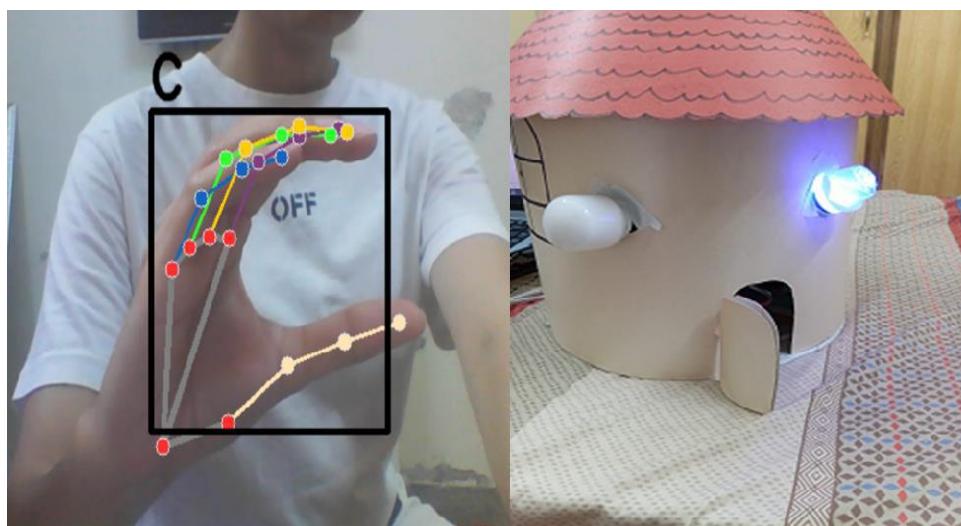


Figure 9: ASL Gesture "C" makes white bulb off

To turn off the lights, a hand motion labeled 'C' was required as indicated in Figure 9. This action is facilitated by a camera in the system, which is capable of detecting white lights, as shown in Figure 9. This illustrates the system's ability to perform multi-stage tasks using the corresponding ASL hand motions.



Figure 10: Hand Gesture dynamics

The system was also evaluated for its ability to accurately recognize hand gestures in dimly lit settings with varying light sources, as shown in Figure 10. The system successfully recognized and tracked multiple hands pointing in different directions, allowing for interaction with multiple individuals or performing complex hand activities involving coordination

among all hands. The tests demonstrated the system's capability to detect hand gestures both when hands are in a relaxed position and when standing.

The system produced favorable results under various lighting conditions, maintaining proper functionality even with reduced light levels. Additionally, the system's performance remained consistent despite changes in the distance between the hand and the camera. The CNN model was trained using 4,500 samples for each gesture. A combination of Adam (Adaptive Moment Estimation) and SGD (Stochastic gradient descent) optimizers was employed to enhance model performance. Adam was selected for its ability to converge quickly, while SGD was used to achieve more optimal solutions. The high variance limitation in SGD is mitigated by Adam, resulting in both fast convergence and solution optimality. The probability of gestural detection was assessed using tables 1 and table 2 under various asymmetric lighting conditions.

Table 3: Probability of Gesture Detection in Multiple Lighting Conditions

Lightning Condition	Low	Medium	High
N(A)	3645	3974	4320
N(S)	4500	4500	4500
P(A)= N(A)/N(S)	0.81	0.88	0.96

Table 4: Probability of Gesture Detection with Different Distances

Varying Distance	1m	2m	4m
N(A)	4230	4095	3735
N(S)	4500	4500	4500
P(A)= N(A)/N(S)	0.94	0.91	0.83

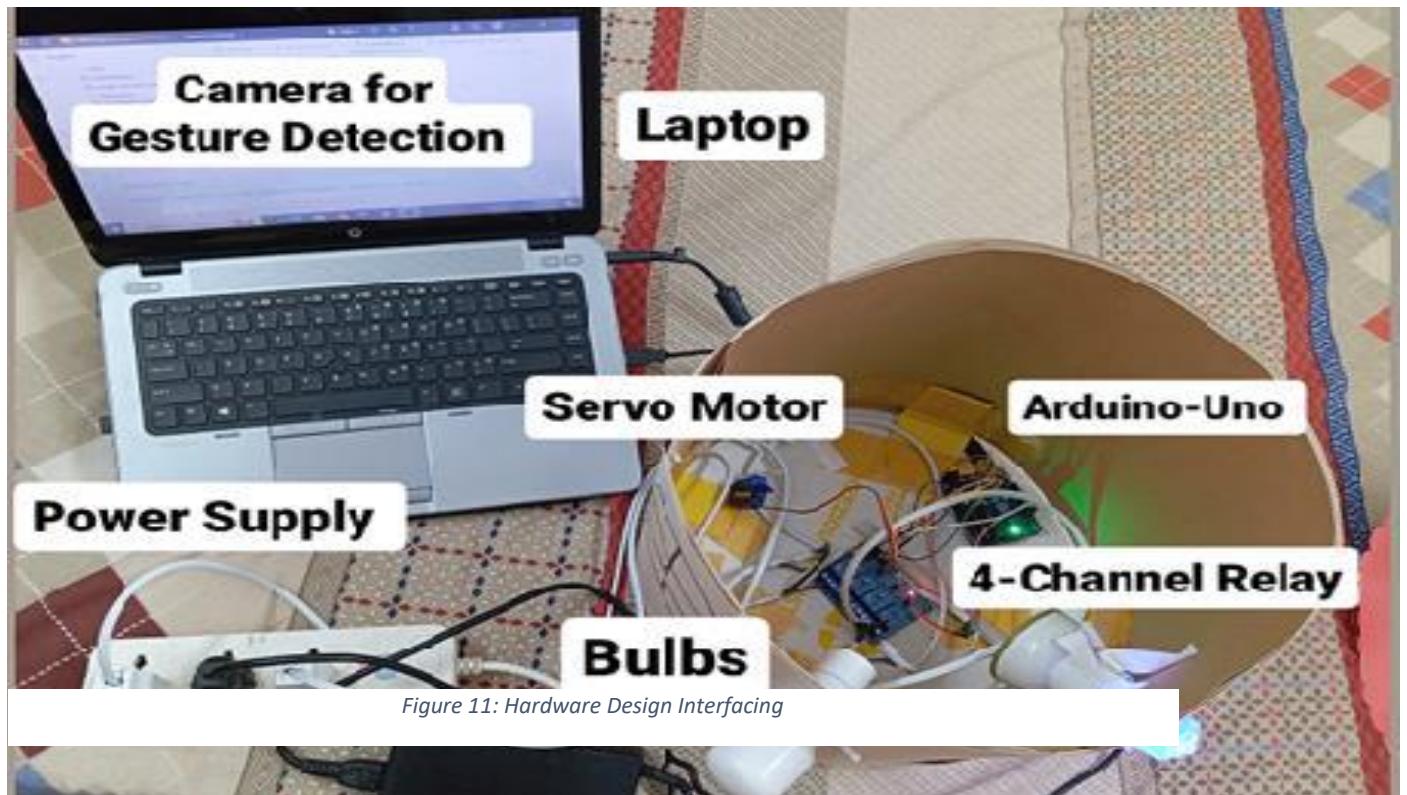


Figure 11: Hardware Design Interfacing

Using Arduino Uno microcontrollers connected to the laptop via COM4, interaction between the software and hardware was established. The Arduino controls a relay module and a servo motor to execute commands. The servo motor,

connected to one of the digital pins on the Arduino, adjusts to specific angles based on hand gestures detected by the system's gesture control automation.

Additionally, a four-channel relay module connected to the Arduino manages the power to the light bulbs according to the software commands. All devices are connected to a common ground (GND) and a 5V power supply rail provided by the microcontroller. The system includes two 0-watt bulbs, which are LED or low-power bulbs used as visual indicators or actuators, lighting up in response to specific gesture movements.

4. Conclusion

In this work it is shown how home automation could be improved by incorporating an IoT platform and hand gesture recognition technology. The hand gesture recognition system is quite advanced and can even be integrated with people's mobile phones and is already internet-enabled, so that people can operate household devices monitored and controlled through the Internet, with mobile phones or voice assistants, using IoT platforms like MQTT or CoAP. This gesture interface is supplemented with a hand landmarks tracking model that allows controlling devices intuitively and naturally by performing simple gestures to the camera. Gesture controlled lights, kitchen appliances, and even door servos are made possible through the attaching of Arduino-UNO microcontroller to relays and precise gestures. All of these contribute to better user engagement with technology, thus making home automation systems put to their maximum benefits. It is however possible to extend the functionality of the system to other areas such as improving the accessibility of people, care for the elderly or responding to emergencies, which proves that considerable progress has been made in the interaction between a person and a computer and home technology.

Acknowledgment

We are thankful to the research reviewers who helped in writing and improving these guidelines.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

5. References

- [1]. Yang, LI and Huang, Jin and Feng, TIAN and Hong-An, WANG and Guo-Zhong, DAI, "Gesture interaction in virtual reality," *Virtual Reality & Intelligent Hardware*, pp. 84--112, 2019.
- [2]. Kavakli, Manolya, "Gesture recognition in virtual reality," *International Journal of Arts and Technology*, vol. 1, no. 2, pp. 215--229, 2008.
- [3]. Sagayam, K Martin and Hemanth, D Jude, "Hand posture and gesture recognition techniques for virtual reality applications: a survey," *Virtual Reality*, vol. 21, pp. 91--107, 2017.
- [4]. Zhang, Qi and Zhu, Wenzhe and Zhu, Qing, "Real world hand gesture interaction in virtual reality," *Journal of physics: Conference series*, s.l. : IOP Publishing, 2019, p. 012027.
- [5]. Kurian, Biya and Regi, Jerom and John, Dennis and Hari, P and Mahesh, Therese Yamuna, "Visual Gesture-Based Home Automation," *IEEE*, pp. 286--290, 2023.
- [6]. Lippmann, Gabriel, "La photographie integrale," *Comptes-Rendus*, vol. 146, pp. 446--451, 1908.
- [7]. Davis, James and Shah, Mubarak, "Recognizing hand gestures," *Computer Vision—ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6, 1994 Proceedings, Volume 1 3*, s.l. : Springer, 1994, pp. 331--340.
- [8]. Wu, Ying and Huang, Thomas S, "Hand modeling, analysis and recognition," *IEEE Signal Processing Magazine*, vol. 18, no. 3, pp. 51--60, 2001.
- [9]. Park, Gyutae and Chandrasegar, Vasantha Kumar and Koh, Jinhwan, "Accuracy enhancement of hand gesture recognition using CNN," *IEEE Access*, vol. 11, pp. 26496--26501, 2023.
- [10]. Sun, Jing-Hao and Ji, Ting-Ting and Zhang, Shu-Bin and Yang, Jia-Kui and Ji, Guang-Rong, "Research on the hand gesture recognition based on deep learning," *2018 12th International symposium on antennas, propagation and EM theory (ISAPE)*, s.l. : IEEE, 2018, pp. 1--4.
- [11]. Licsar, Attila and Sziranyi Tamas, "Dynamic training of hand gesture recognition system," *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, s.l. : IEEE, 2004, pp. 971--974.
- [12]. Hendricks, Drew, "The History of Smart Homes," *IoT Evolution World*, April 22, 2014. <https://www.iotevolutionworld.com/m2m/articles/376816-history-smart-homes.htm>.
- [13]. Varriale, Luisa and Briganti, Paola and Mele, Stefania, "Disability and home automation: insights and challenges within organizational settings," *Springer*, pp. 47--66, 2020.

- [14]. Asadullah, Muhammad and Raza, Ahsan, "An overview of home automation systems," *2016 2nd international conference on robotics and artificial intelligence (ICRAI)*, s.l. : IEEE, 2016, pp. 27–31.
- [15]. Kim, Youngwook and Toomajian, Brian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.
- [16]. Katre, Sharda R and Rojatkar, Dinesh V, "Home automation: past, present and future," *International research journal of engineering and technology*, vol. 4, no. 10, pp. 343–346, 2017.
- [17]. A. Ahmed, P. Botsinis, S. Won, L. -L. Yang and L. Hanzo, "EXIT Chart Aided Convergence Analysis of Recursive Soft mm-Sequence Initial Acquisition in Nakagami-m Fading Channels," in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4655-4660, May 2018, doi: 10.1109/TVT.2018.2789912.
- [18]. F. Che, A. Ahmed, Q. Z. Ahmed, S. A. R. Zaidi and M. Z. Shakir, "Machine Learning Based Approach for Indoor Localization Using Ultra-Wide Bandwidth (UWB) System for Industrial Internet of Things (IIoT)," *2020 International Conference on UK-China Emerging Technologies (UCET)*, Glasgow, UK, 2020, pp. 1-4, doi: 10.1109/UCET51115.2020.9205352.
- [19]. Mushtaq, Bilal, and Sohail Khalid. "Design of miniaturized single and dual-band bandpass filters using diamond-shaped coupled line resonator for next-generation wireless systems." *International Journal of Microwave and Wireless Technologies* 15, no. 3 (2023): 375-383. doi: <https://doi.org/10.1017/S1759078722001416>
- [20]. B. Mushtaq, S. Khalid and M. A. Rehman, "Design of a Compact Novel Stub Loaded Pentaband Bandpass Filter for Next Generation Wireless RF Front Ends," in *IEEE Access*, vol. 10, pp. 109919-109924, 2022, doi: 10.1109/ACCESS.2022.3214313.
- [21]. B. Mushtaq, M. A. Rehman, S. Khalid and M. Alhaisoni, "Design of Tri-Band Bandpass Filter Using Modified X-Shaped Structure for IoT-Based Wireless Applications," in *IEEE Embedded Systems Letters*, vol. 16, no. 2, pp. 194-197, June 2024, doi: 10.1109/LES.2023.3325898.
- [22]. B. Mushtaq, M. A. Rehman, A. Hussain and M. J. Abbass, "A Highly Selective Dual Bandpass Filter using Couple Line Resonator for Modern Wireless Communication Systems," *2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, Pakistan, 2023, pp. 1-5, doi: 10.1109/iCoMET57998.2023.10099212.
- [23]. Abdulla, Abdulrahman Ihsan and Abdulraheem, Ahmad Sinali and Salih, Azar Abid and Sadeeq, MA and Ahmed, Abdulraheem Jamel and Ferzor, Barwar M and Sardar, Omar Salih and Mohammed, Sarkaft Ibrahim, "Internet of things and smart home security," *Technol. Rep. Kansai Univ*, vol. 62, no. 5, pp. 2465–2476, 2020.
- [24]. UN DESA. 2023. The Sustainable Development Goals Report 2023: Special Edition - July 2023. New York, USA: UN DESA. © UN DESA. <https://unstats.un.org/sdgs/report/2023/>



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:
Majid et al. (2024).
"Classification of Gastric Infections Using Deep Learning Features". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024 Majid et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

Classification of Gastric Infections Using Deep Learning Features

Abdul Majid ^{a*}

Mehwish Zafar ^a

Fadia Ali Khan ^a

^a Department of Computer Science, HITEC University Museum Road, Taxila, Pakistan.

abdul.majid@hitecuni.edu.pk

mehwish.zafar@hitecuni.edu.pk

fadia.ali@hitecuni.edu.pk

* Corresponding author: abdul.majid@hitecuni.edu.pk

Abstract:

Stomach diseases are the deadliest and the major cause of human mortality in the world. Due to the harmful effects of stomach lesions, these diseases need to be diagnosed at an early stage. Artificial intelligence-based automated models can help to recognize stomach diseases through wireless capsule endoscopy (WCE) images, and the mortality rate can be reduced significantly. This article presents a new deep learning-based model to recognize stomach diseases in WCE images. The significance of the presented classification model is to enhance the precision, efficacy, and timely diagnosis of gastric diseases. The proposed approach has the potential to assist healthcare practitioners in making decisions for treatment, and facilitating medical investigations, ultimately resulting in improved patient care and outcomes. Initially, a convolutional neural network (CNN) architecture is developed containing a total of 45 layers. The proposed network includes 14 convolution, 4 average pooling, 3 batch normalization (BN), 3 fully connected (FC) layers, and one dropout layer. Each convolution is followed by a ReLU layer, and a total of 16 ReLU layers are utilized. In this architecture, 10 layers are utilized in parallel including four convolutional, four ReLU, and two batch normalization layers. Then the proposed CNN architecture is trained on the five classes of the HyperKvasir dataset. Following the model training, deep features are collected from the two FC layers. The derived features from both FC layers are then concatenated, and the final feature set is forwarded to the multiple machine learning (ML) classifiers for the final prediction. After extensive experiments, a maximum classification accuracy of 98.90% is obtained. Finally, a comparison of the proposed technique is carried out with the latest existing methods that show a significant performance.

Keywords: deep learning, feature extraction, feature fusion, stomach diseases, classification

1. Introduction

Gastrointestinal (GI) disease recognition is an active research area in medical imaging. Different types of GI diseases include ulcers, polyps, esophagitis, and bleeding. The most common disease is ulcer among all lesions. These GI lesions cause major human mortalities [1, 2]. In 2021, the newly reported cases of gastric cancer in the United States were 26,560. Among them, male patients were 16,160 and females were 10,400. The total approximate mortalities were 11,180, of which 6740, and 4440 in males and females respectively [3]. As per the American Cancer Society, 26,380 fresh instances of stomach cancer estimated cases, there would be 15,900, and 10,480 in men, and women, respectively. The total estimated mortalities were 11,090 which include 6690 and 4400 in males and females, respectively [4].

The wireless capsule endoscopy (WCE) is utilized to observe the GI tract which is a medical imaging procedure. This process utilizes a tiny camera for capturing videos of the GI tract. The identification of diseases from the WCE videos or frames is a hectic and time-consuming task for healthcare professionals. Nearly 50,000 images are generated during the WCE procedure. Selection of abnormal frames is also a challenging task because only 5% of images are lesion-affected [5]. The development of automated computer-aided diagnostic (CAD) models can resolve these issues. The CAD systems collect the information from WCE frames for disease detection. The fundamental steps of a CAD system comprise pre-processing, feature extraction, and recognition.

Early diagnosis of stomach problems may lower death rates. Because human lesion identification is time-consuming, automated CAD models may be used for early illness detection. Computer vision (CV) scientists have created automated CAD systems for gastric abnormalities prediction. These automated systems utilize handcrafted and deep learning (DL) features and predict diseases using supervised learning techniques. In the existing methodologies, the researchers have used point features [6], color features from different transformations [7], and texture attributes [8, 9]. Additionally, the convolutional neural network (CNN) features [10-14] have also been utilized from pre-trained deep models.

DL is now essential for medical image identification and segmentation. DL is an influential machine learning (ML) method for automatically categorizing medical lesions [15-17]. It aids health professionals in disease identification [18]. Deep learning CNNs derive gastric image features. CNNs can handle signals, pictures, videos, and multidimensional data. CNN's early layers supply low-level characteristics, whereas its later layers provide high-level information. The key layers present in a CNN model include various layers like convolution, pooling, fully connected, and classification output. On all image pixels convolution operation is applied by the convolutional layers. FC layers vectorize retrieved information for illness recognition [19].

1.1. Research Objectives

The focus of this work is to develop and design effective techniques for gastric disease recognition that help to diagnose the disorders at an early stage and help to save lives with the help of advanced deep and machine learning approaches. The use of CAD systems can help to increase the survivability of the patients by detecting the disease at the early stages. Moreover, the aim is to enhance accuracy by reducing the rate of false positives and negatives and provide the early recognition of disease by overcoming the challenges of noise, image quality, and similarities among the infected lesions which may lead to false categorization.

1.2. Research Contributions

A new deep CNN-based technique is designed and implemented in this paper for the classification of stomach abnormalities. The prominent characteristics of the proposed model include:

- A 45-layered CNN architecture is developed introducing 10 parallel layers.
- After training the deep CNN model, the features are extracted and fused. Then the classification is performed on the final feature set using multiple ML classifiers.

The other sections of this article include existing work, proposed technique, results and discussion, and conclusion.

2. Related Work

During the last few years, extensive work has been done to design stomach disease identification systems. These systems are based on artificial intelligence-based methodologies such as CV, ML, DL, and other algorithmic techniques [20-22]. Previous studies have shown promising outcomes in terms of accurate disease identification when compared to manual diagnostic methods. In this section, some latest techniques are discussed for the recognition of GI tract abnormalities. Researchers [7], combined deep CNN features with multiple classical features for the recognition of stomach lesions. In the deep features extraction step, a pre-trained model VGG16 [23] was utilized. For handcrafted features, researchers extract discrete cosine, discrete wavelet transform, and color features. After the fusion of all extracted features, the fused feature vector was forwarded to the optimization method namely genetic algorithm (GA) to select the robust features. The highest recognition rate of 96.5% was achieved on the ensemble classifier. Authors introduced a CNN model [24] that utilizes a spatial attention method, including both encoder and decoder layers, for the categorization of GI abnormalities. The computed results are 93.19% of accuracy, precision 92.8%, recall 92.7%, and F1-score 92.8%. The study [25] revealed a technique to segment and classify stomach lesions using WCE images. In the segmentation phase, researchers introduced deep labv3 followed by the pre-trained ResNet-50 architecture. The trained model performed accurate pixel-wise classification. After segmenting the lesion area, the proposed classification method extracted features from the pre-trained ResNet-50 CNN architecture and achieved a prediction rate of 98%. In [26], the authors conducted a comprehensive series of tests to acquire the results for the identification of GI abnormalities. The researchers selected five well-known pre-trained CNN. These models were then subjected to re-training on the Kvasir database. The achieved recognition accuracies XceptionNet 98.2%, Inception-v3 90%, MobileNet 97.6%, ResNet 92.3%, and VGG-16 98.3%. The results indicate that the VGG-16 model obtained a maximum recognition rate of 98.3%.

A previous study [27] introduced a framework that used DL features to classify stomach lesions. The system employs the VGGNet and InceptionNet architectures to perform deep feature computing. Following the feature extraction from deep networks, the feature vectors were combined and allocated to several machine-learning predictors for illness detection. The machine learning technique Support Vector Machine (SVM) obtained the best accuracy in illness

detection, with an estimated accuracy rate of 98%. Sutton et al. [28] utilized the HyperKvasir dataset for the recognition of ulcerative colitis in endoscopic images. They compared several CNN architectures for the recognition of lesions in the selected database. After extensive experiments, the best accuracy of 87.5%, and 90% of the area under the curve was calculated on the DenseNet-121 model. A convolutional-capsule network [29] was introduced to classify the GI images. The suggested network is a two-stage sample classification technique that effectively utilizes the benefits of mid-level features from CNN architecture and the capsule architecture. The deep attributes were collected from the network and classification experiments were performed for comparison with the other existing methods. The proposed model achieved a 94.83% recognition rate on the Kvasir-V2, and 85.99% on the HyperKvasir dataset. In another study [30], researchers used a bottom-up multi-scale approach to compute low-level features, and the input features were corrected by top-down modules to facilitate attention learning. The use of the backbone stream was employed for the computation of high-level features. A correction loss function was also introduced to validate the combination of both types of features. The proposed system was assessed on the Kvasir dataset and achieved a maximum prediction rate of 97.33%.

3. The Research Problem

In literature, DL models were utilized for the classification of gastric infections using the numerous available procedures. The accuracy was improved by a few of those who merged deep models, but the computational time was much increased. Furthermore, some authors have put forth a combination of classical and deep learning-based techniques, but problems including feature redundancy, computational expense, and model robustness still exist. Therefore, in this work, the aforementioned challenges are considered, and a classification methodology is developed using deep learning features.

4. Proposed Solution

This section comprises a comprehensive representation of the designed approach for the classification of gastric infections. Figure 1 portrays the designed architecture. The major steps described in this section include data augmentation, the development of a new CNN architecture, training of the proposed CNN model, feature extraction, feature fusion, and finally classification.

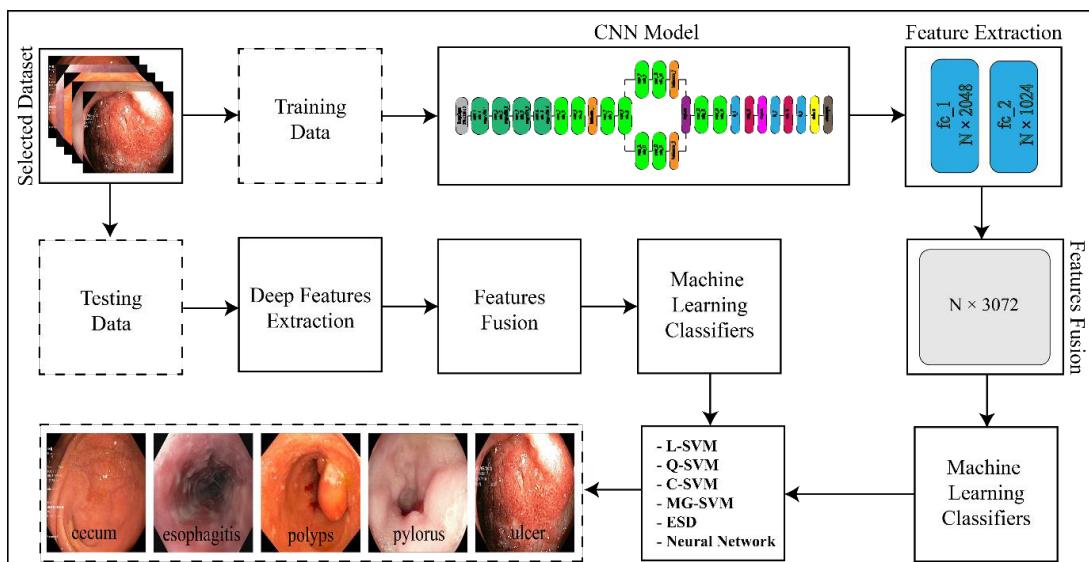


Figure 1: Proposed framework for the classification of gastric abnormalities.

Firstly, augmentation is performed on the selected dataset. In the selected dataset five classes including cecum, esophagitis, polyps, pylorus, and ulcerative colitis contain 1009, 663, 1028, 999, and 851 images, respectively. After applying different image rotation techniques like left to right, and up to down data in each class is increased to 2000 images. After this augmentation process, the dataset consists of 10,000 samples. In this research, a new CNN architecture is developed comprising 45 layers. The developed model is a directed acyclic graph (DAG) network. In the DAG network, layers are arranged in a directed acyclic graph manner. The developed CNN model comprises 14 convolutional, 16 ReLU, 4 average pooling, 3 batch normalization, a dropout, and 3 FC layers. The functional parameters and structure of this CNN model are represented in Figure 2. The functional parameters include layer type, activations, bias (bs), and weights (wgt). Here, a layer-wise brief demonstration of the layer-wise architecture is presented below:

The image input layer is the first layer among all CNN layers. In the confer model, the image input layer of size $256 \times 256 \times 3$ is utilized. The foremost element of CNN is the convolutional layer. It contains a group of filters, and each

filter convolves with the image to obtain the activation map. Here a total of 14 convolution layers are utilized, from which four layers are used in a parallel way. Mathematically, the convolution operation can be depicted as:

$$\text{Output}(i,j) = \sum_{y=0}^{\text{columns}} \sum_{x=0}^{\text{rows}} \text{Filter}(x,y) \cdot \text{Input}(i-x, j-y) \quad (1)$$

A total of 16 ReLU layers are used in this architecture. 14 layers after convolutional layers, and two layers after FC layers. In the ReLU layer, a non-linear activation function is applied and zero replaces the negative values. The average Pooling layer performs a pooling operation on a feature map that computes the average value for a receptive patch. In this architecture, four average pooling layers such as avgpool2d_1, avgpool2d_2, avgpool2d_3, and avgpool2d_4 applied after the conv1, conv2, conv3, and conv4 layers, apart.

The batch normalization (BN) approach is employed to normalize the previous layer's output. It is employed to make learning easy, boost the training procedure, and also help to elude overfitting problems. In the proposed CNN architecture, the first batch normalization layer batchnorm_1 is employed after the conv_6 convolutional layer. The second and third batch normalization layers are utilized after the two parallel relu_10, and relu_12 layers, respectively.

In the proposed architecture, the input is split up into two directions after the relu_8 layer, and 10 parallel layers are introduced in the network. These parallel layers comprise four convolutional layers, four ReLU layers, and two batch normalization layers. The two inputs are concatenated using the depth concatenation layer, and the output is given to the conv_13 layer. The dropout layer of 50% rate is added after the relu_15 layer in the proposed architecture. It means that 50% of neurons are switched off. In this CNN model, three Fully Connected layers are utilized including fc1, fc2, and fc3. The activations on these FC layers fc1, fc2, and fc3 are $1 \times 1 \times 2048$, $1 \times 1 \times 1024$, and $1 \times 1 \times 5$, respectively.

Latterly, the softmax classifier is utilized which implements the softmax function [31]. This function is a generic form of the binary logistic regression function. The mathematical modeling of the mapping function h used in this softmax layer can be shown as:

$$h(m_i, W) = Wm_i \quad (2)$$

Here, the input set is denoted by m , and W represents the weight matrix.

The class output is the final or last layer in a deep CNN architecture. This layer computes cross-entropy loss to classify mutually incompatible classes.

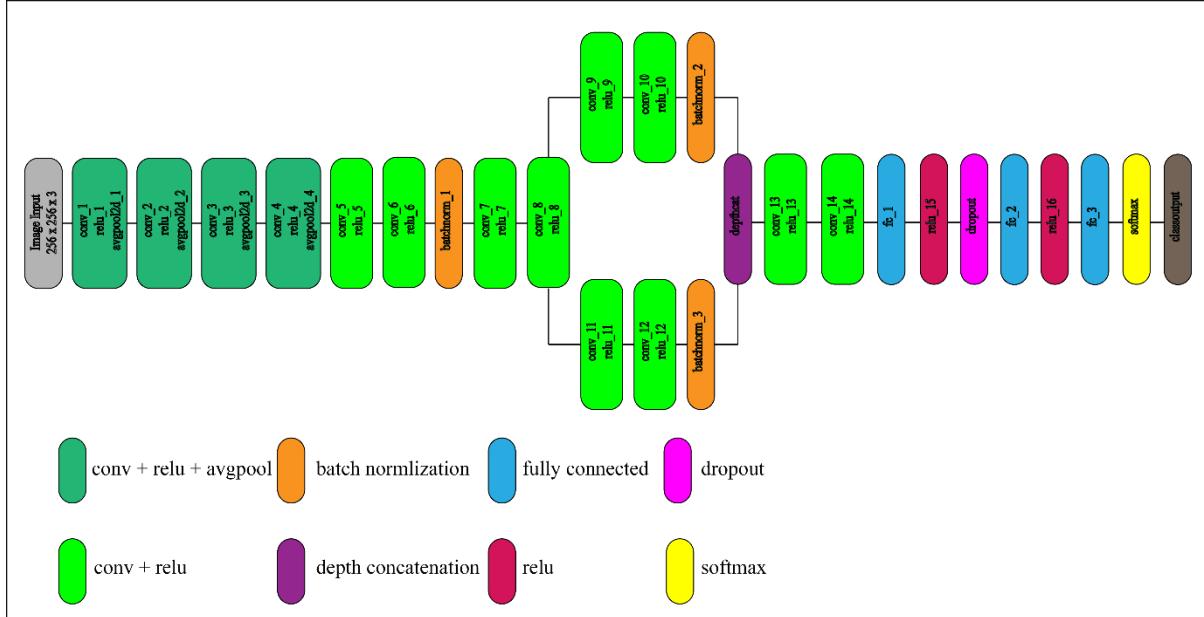


Figure 2: Architecture of the proposed 45-layered CNN model.

The intended CNN architecture is trained on the selected classes of the HyperKvasir dataset. Thus, the training process is completed in 16 epochs, and each epoch runs in 109 iterations. The learning rate is set as constant with a value of 0.01. The size of the minibatch is adjusted to 64, and stochastic gradient descent with momentum (SGDM) is utilized as the network optimizer. A 70:30 data split approach is utilized in the proposed architecture, 70% for training, and 30% for validation. The data for training and validation is randomly split after each epoch. After the completion of all iterations, we achieved a validation accuracy of 95.77%.

After the model training, deep attributes are obtained from the fc1 and fc2 layers. The feature sets of dimensions $N \times 2048$, and $N \times 1024$ collected from fc_1, and fc_2 layers respectively. Here, N is denoting the entire training samples.

Both feature vectors $N \times 2048$, and $N \times 1024$ are transformed into a single vector by serial concatenation, thus $N \times 3072$ is the size of the final vector. Whereas the fused vector is fed to various classifiers to classify infections. The mathematical representation of the final vector can be formulated as:

$$(fused)_{N \times 3072} = U[(fc_1)_{N \times 2048}, (fc_2)_{N \times 1024}] \quad (3)$$

5. Experiments

The implementation and experiments of the proposed methodology are carried out on MATLAB R2021a using the desktop computer having specs core i5 8th generation, 16 GB RAM, and 4 GB graphics memory. In this article, all experimentations are conducted on described ratios of training and testing data. The results are collected in four experiments using 50:50, 60:40, 70:30, and 80:20 dataset training and testing ratios, respectively. The proposed methodology is tested on multiple classifiers and six high-performance classifiers are selected for further experiments. The selected classifiers are Linear Support Vector Machine (L-SVM), Cubic Support Vector Machine (C-SVM), Quadratic Support Vector Machine (Q-SVM), Medium Gaussian Support Vector Machine (MG-SVM), Ensemble Subspace Discriminant (ESD), and neural network. Different assessment measures such as accuracy (Acc), F1-score (F1-S), recall (Rc), precision (Prc), and computational time (seconds) are utilized to analyze the functionality of classifiers.

5.1. Experiment 1 Using 50:50 Training and Testing Data

In the first experiment, 50:50 data split ratios are used for both training and testing and all the outcomes are given in Table 1. This experiment observed a maximum prediction rate of 98.40% on the L-SVM predictor. The accuracy can be validated in Figure 3. In comparison to the other predictors in the experiment, the L-SVM classifier took the shortest time of 30.467 seconds to train. The extra performance metrics obtained on the L-SVM classifier are 98.20% precision, 98.40% recall, and 98.40% F1-score. The accuracies achieved on Q-SVM, C-SVM, MG-SVM, ESD, and neural network classifiers are 98.20%, 98.30%, 98.00%, 98.30%, and 97.90%, respectively. The highest training time is 726.55 seconds calculated on the ESD classifier.

Table 1: Experiment 1 results for stomach disease classification.

Classifier	Acc %	Prc %	Rc %	F1-S %	Time (seconds)
L-SVM	98.40	98.20	98.40	98.40	30.467
Q-SVM	98.20	98.20	98.60	98.40	136.81
C-SVM	98.30	98.10	98.50	98.30	117.18
MG-SVM	98.00	98.10	98.30	98.30	140.47
W-KNN	97.90	98.00	98.20	98.20	239.72
ESD	98.30	98.20	98.20	98.20	726.55

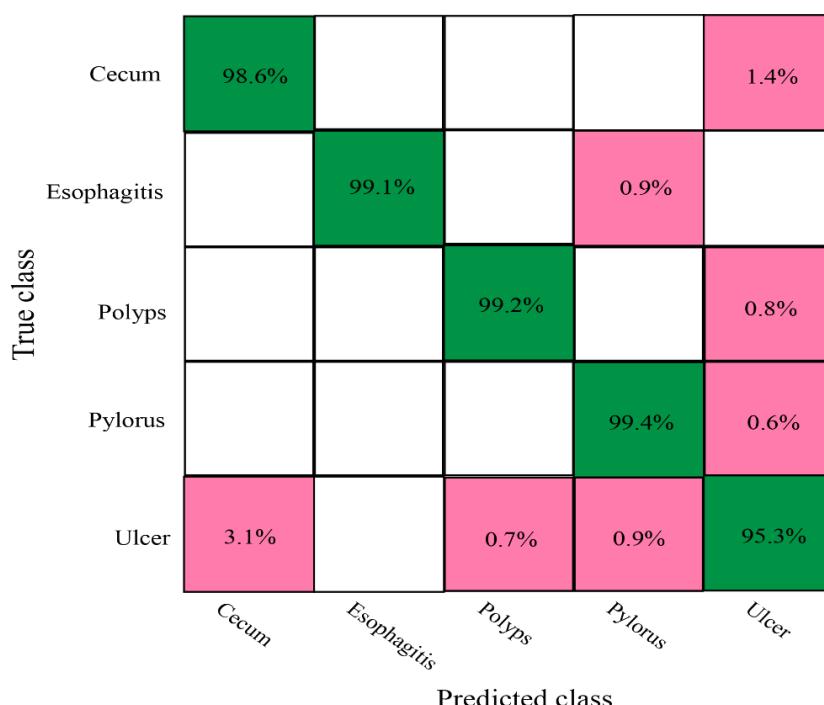


Figure 3: L-SVM classifier's confusion matrix for experiment 1.

5.2. Experiment 2 Using 60:40 Training and Testing Data

In experiment 2, 60% of the selected data is utilized for training and 40% data for testing purposes. Table 2 depicts the overall results of different classifiers. The ESD classifier attains the highest accuracy of 98.60% in this experiment, but it takes the highest computational time of 597.80 seconds among all the classifiers. The other measures calculated on the ESD classifier are a precision of 98.60%, 98.60% recall, and an F1-score of 98.60%. The working of the ESD classifier can be validated in Figure 4. The accuracy performance of other classifiers is 98.40%, 98.40%, 98.50%, 98.40%, and 98.50% for L-SVM, Q-SVM, C-SVM, MG-SVM, and neural network, respectively. The lowest training time of 25.235 seconds is observed on the neural network classifier.

Table 2: Experiment 2 results for stomach disease classification.

Classifier	Acc %	Prc %	Rc %	F1-S %	Time (seconds)
L-SVM	98.40	98.20	98.40	98.40	35.109
Q-SVM	98.40	98.20	98.40	98.40	36.953
C-SVM	98.50	98.20	98.60	98.60	37.796
MG-SVM	98.40	98.20	98.20	98.20	44.018
Neural Network	98.50	98.60	98.60	98.60	25.235
ESD	98.6	98.6	98.6	98.6	597.80

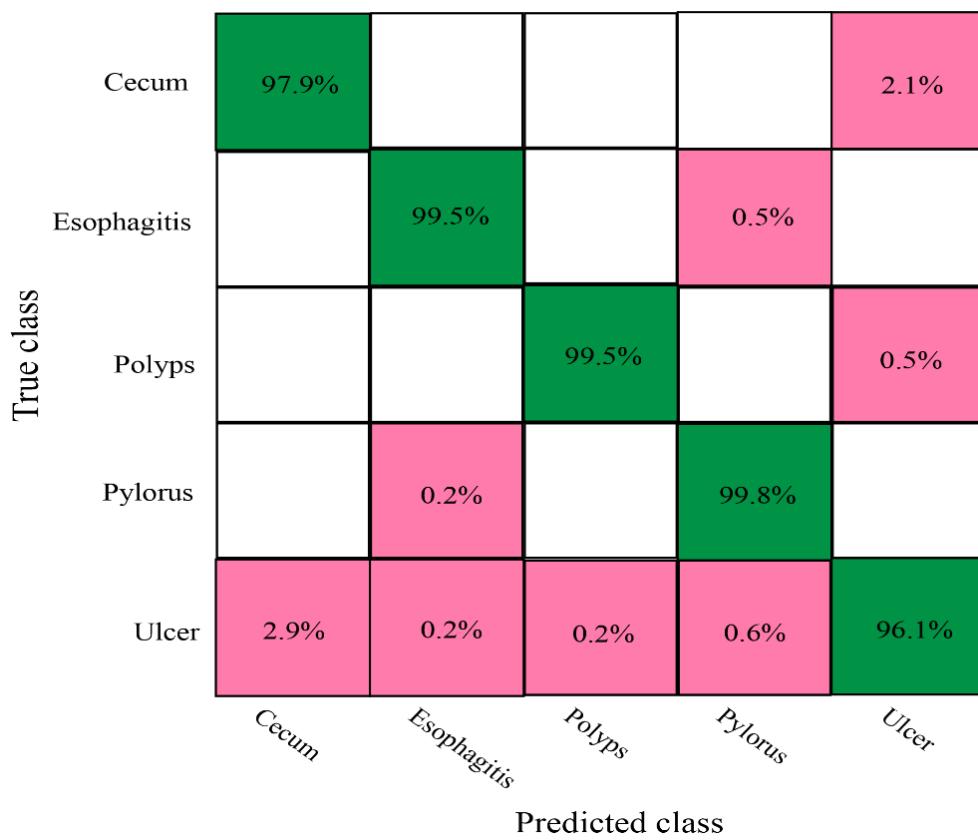


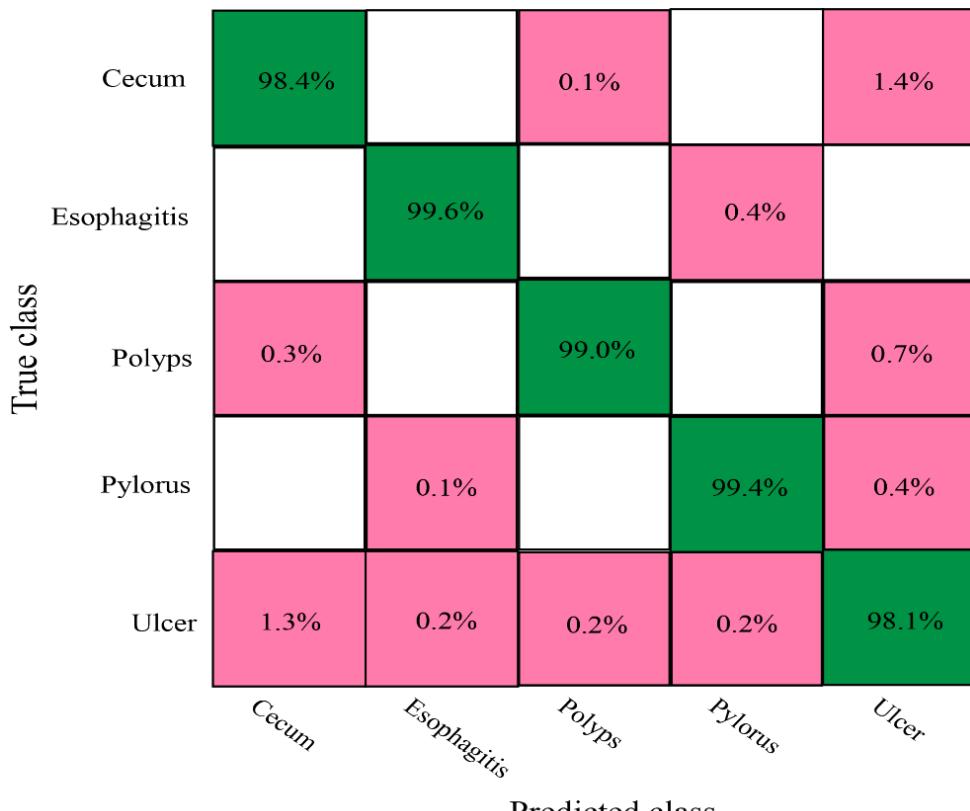
Figure 4: ESD classifier's confusion matrix for experiment 2.

5.3. Experiment 3 Using 70:30 Training and Testing Data

Experiment 3 is conducted using the 70:30 data split approach for the training and testing for the computation of results. Maximum accuracy of 98.90% is obtained on three classifiers Q-SVM, C-SVM, and ESD as depicted in Table 3. For the Q-SVM classifier, precision, recall, and F1-score are 98.90%, 99.10%, and 98.90%. As well as Figure 5 depicts the outcomes of the Q-SVM classifier. Results computed on the C-SVM are 98.90% for precision, 99.00% for recall, and 98.60% for F1-score. The ESD classifier attained 98.90% for precision, 98.90% for recall, and an F1 score of 98.90%. The computational time for Q-SVM, C-SVM, and ESD classifiers is 42.246 seconds, 42.284 seconds, and 554.80 seconds, respectively. The computational time of 24.733 seconds is the best among all and was observed on the neural network classifier with 98.40% accuracy. The accuracy computed on L-SVM, and MG-SVM is 98.50% for both classifiers.

Table 3: Experiment 3 results for stomach disease classification.

Classifier	Acc %	Prc %	Rc %	F1-S %	Time (seconds)
L-SVM	98.50	98.50	98.70	98.70	434.33
Q-SVM	98.90	98.90	99.10	98.90	42.246
C-SVM	98.90	98.90	99.00	98.60	42.284
MG-SVM	98.50	98.50	98.50	98.50	170.98
Neural Network	98.40	98.70	98.70	98.70	24.733
ESD	98.90	98.90	98.90	98.90	554.80

**Figure 5:** Q-SVM classifier's confusion matrix for experiment 3.

5.4. Experiment 4 Using 80:20 Training and Testing Data

Experiment 4 is performed using the 80:20 approach for 80% training and 20% testing, and the results are portrayed in Table 4. During this experimentation, the maximum recognition rate of 98.50% is observed on Q-SVM, and C-SVM having computational times of 47.717 seconds, and 48.405 seconds, respectively. The observed performance on the Q-SVM is precision, recall, and F1-score which are 98.60%, 98.80%, and 98.60% individually, and can be validated in Figure 6. The precision of 98.40%, recall of 98.80%, and F1-score of 98.40% were recorded for C-SVM respectively. The accuracies for L-SVM, MG-SVM, ESD, and neural network are 98.40%, 98.20%, 98.40%, and 98.40%, respectively.

Table 4: Experiment 4 results for stomach disease classification.

Classifier	Acc %	Prc %	Rc %	F1-S %	Time (seconds)
L-SVM	98.40	98.20	98.60	98.40	41.084
Q-SVM	98.50	98.60	98.80	98.60	47.717
C-SVM	98.50	98.40	98.80	98.40	48.405
MG-SVM	98.20	98.20	98.40	98.20	57.645
Neural Netwrok	98.40	98.60	98.60	98.40	33.746
ESD	98.40	98.40	98.60	98.20	657.85

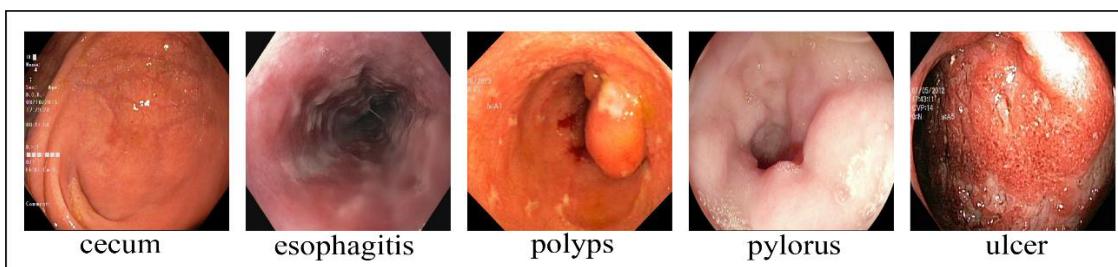
	Cecum		0.1%		1.4%
True class		99.6%		0.4%	
	0.2%		98.9%	0.1%	0.7%
		0.1%		99.2%	0.6%
True class	2.3%	0.1%	0.3%	0.9%	96.3%
Predicted class	Cecum	Esophagitis	Polyps	Pylorus	Ulcer

Figure 6: Q-SVM classifier's confusion matrix for experiment 4.

In experiment 4, a higher computational time is observed in contrast to the other three experimentations. The maximum computational time of 657.85 seconds is calculated on the ESD classification algorithm. The neural network classifier is trained in 33.746 seconds, which is the best computational time among all classifiers.

6. Results and Discussion

Here, a comprehensive discussion of experiments and outcomes is presented. To evaluate the developed model, the HyperKvasir [32] dataset is used. This is an open-access dataset collected from 2008 to 2016 at Bærum Hospital. HyperKvasir mainly contains four sets of records. The complete dataset comprises a total of 110,079 samples. The labeled image record is considered to perform the experiments. The total labeled images of HyperKvasir contains 10,662 images. In the experiments, five classes are utilized to collect the results. These five classes include cecum, esophagitis, polyps, pylorus, and ulcerative colitis. The images of the selected dataset are presented in Figure 7 as the sample. The dataset classes cecum, esophagitis, polyps, and pylorus contain 1009, 663, 1028, and 999 images, respectively. The ulcerative colitis images were graded using the Mayo endoscopic sub-score. The images are categorized into grades (G) 0 to 1, G 1 to 2, G 2 to 3, and G 3 which consist of 35, 201, 11, 443, 28, and 133 images, respectively. We grouped all grading categories into a single class of ulcerative colitis that contains 851 images. After applying augmentation techniques, the data in all five classes is increased. After this, there are 2000 images per class.

**Figure 7:** Sample images of the selected classes of the HyperKvasir dataset.

This research introduced a deep CNN architecture for the prediction of stomach infections and is evaluated using the five classes of the HyperKvasir dataset. After training the proposed CNN architecture, features are computed and merged into one vector, and four experiments are conducted on different ratios for training and testing data. In the first experiment, using the 50:50 data, the maximum 98.40% recognition rate is obtained on the L-SVM classifier. The minimum computational time is also observed on this classifier which is 30.467 seconds. The second experiment using the 60:40 training and testing date observed the highest recognition rate of 98.60% on the ESD classifier. But the calculated time on ESD is the worst among all classifiers which is 597.80 seconds. The neural network classifier trained in the best time of 25.235 seconds with 98.40% accuracy. In the third experiment, the best accuracy of 98.90% is obtained on three classifiers, namely Q-SVM, C-SVM, and ESD. The less training time of 24.733 seconds took neural

network classifier. Moreover, the accuracy observed on the neural network classifier is worse in contrast to others. The fourth experiment computed the greatest prediction rate of 98.50% on Q-SVM, and C-SVM with the computational time of 47.717 seconds, and 48.405 seconds, respectively. The MG-SVM classifier gives a minimum accuracy of 98.2%. The minimum and maximum training time is 33.746 seconds, and 657.85 seconds observed on the neural network, and ESD classifiers, respectively.

Table 5: Accuracy comparison of different training and testing ratios.

Dataset Split Ratio	Classifier	Acc (%)
50:50	L-SVM	98.40
60:40	ESD	98.60
70:30	Q-SVM	98.90
	C-SVM	
	ESD	
80:20	Q-SVM	98.50
	C-SVM	

As per the results, there is no big difference in accuracy using different training and testing ratios. However, there is a slight difference in the computational time of classifiers in all experiments. In the 70:30 data split ratio, the highest classification accuracy of 98.9% is obtained, and the minimum accuracy of 98.00% is observed in the 50:50 data. In Table 5, a comparison of accuracy is presented on different training and testing ratios of the dataset. Moreover, in terms of recognition time, the neural network classifier performed best among all classifiers.

Table 6: Comparison of classification outcomes with the latest existing techniques.

Method	Year	Acc (%)
Lonseko et al. [24]	2021	93.19
Amin et al. [25]	2021	98.00
Dheir et al. [26]	2022	98.30
Haile et al. [27]	2022	98.00
Sutton et al. [28]	2022	87.50
Wang et al. [29]	2022	94.83
Li et al. [30]	2022	97.33
Thomas et al. [33]	2023	98.01
Sivari et al. [34]	2023	98.53
Mukhtorov et al. [35]	2023	98.28
Proposed	2024	98.90

Furthermore, the comparison of results is also carried out with some latest stomach lesion classification techniques presented in Table 6. The contrast demonstrates that the proposed framework outperforms existing techniques and achieves a promising accuracy of 98.90%. The reason behind obtaining higher classification results as compared to existing methods is the addition of parallel layers in the CNN model. Two convolutional layers are utilized in a parallel way, and then after both layers batch normalization layers are introduced which enhances our validation accuracy. As validation accuracy enhanced, the more precise classification accuracy was obtained on different classifiers.

7. Conclusions and Future Direction

A new deep CNN-based framework for the classification of stomach infections has been introduced. A novel CNN model is developed and trained on five classes of the HyperKvasir dataset. After conducting multiple experiments, the maximum recognition rate of 98.90% is observed which confers the promising performance obtained by the fusion of features extracted from the proposed CNN architecture. Extensive experiments were performed during the training of the CNN model by changing the layer position and type to enhance the validation accuracy. Then, in the proposed CNN model, we utilized parallel convolutional and batch normalization layers. The addition of parallel layers enhanced the performance and produced more accurate results. Later, we can work to improve the classification results by reducing CNN parameters and achieving the best accuracies by using minimal robust features and will test on some other datasets to ensure the validity of the architecture. In the real-world applications of gastric infections categorization framework, significant crucial tasks arise from both clinical integration and computational cost. The accurate and high-performance architectures require very expensive hardware resources like GPUs, Memories Graphical cards, etc. which the clinics do not have with limited resource settings. Additionally, these frameworks are energy intensive having long training that makes it difficult in real-time use. Moreover, clinicians may resist the adoption due to trust issues with

models when it comes to critical diagnostic factors so in the future the main concern is to develop an accurate model that resolves all these barriers and becomes helpful in the real world.

Acknowledgment

We are thankful to the research reviewers who helped in writing and improving these guidelines.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

8. References

- [1] Sharif, M., et al., Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence*, 2021. 33(4): p. 577-599.
- [2] H. Ali, et al., "A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 2635-2707, 2020.
- [3] F. Mohammad and M. J. S. Al-Razgan, "Deep Feature Fusion and Optimization-Based Approach for Stomach Disease Classification," 2022, vol. 22, no. 7, pp. 2801.
- [4] K. R. Yabroff, et al., "Association of the COVID-19 pandemic with patterns of statewide cancer services," 2022, vol. 114, no. 6, pp. 907-909.
- [5] M. S. Ayyaz, et al., "Hybrid Deep Learning Model for Endoscopic Lesion Detection and Classification Using Endoscopy Videos," 2022, vol. 12, no. 1, pp. 43.
- [6] E. Tuba, M. Tuba, and R. Jovanovic, "An algorithm for automated segmentation for bleeding detection in endoscopic images," in 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017.
- [7] A. Majid, et al., "Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection," 2020, vol. 83, no. 5, pp. 562-576.
- [8] B. Li and M. Q.-H. J. I. T. o. I. T. i. B. Meng, "Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 323-329, 2012.
- [9] H. Ali, et al., "Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images," *Comput. Methods Programs Biomed.*, vol. 157, pp. 39-47, 2018.
- [10] M. A. Khan, et al., "StomachNet: Optimal deep learning features fusion for stomach abnormalities classification," *IEEE Access*, vol. 8, pp. 197969-197981, 2020.
- [11] J. Amin, et al., "Liver Tumor Localization Based on YOLOv3 and 3D-Semantic Segmentation Using Deep Neural Networks," *Diagnostics*, vol. 12, no. 4, p. 823, 2022.
- [12] M. A. Khan, et al., "Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine," *Comput. Electr. Eng.*, vol. 90, p. 106960, 2021.
- [13] M. A. Khan, et al., "An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 15929-15948, 2020.
- [14] M. A. Khan, et al., "Developed Newton-Raphson based deep features selection framework for skin lesion recognition," *Pattern Recognit. Lett.*, vol. 129, pp. 293-303, 2020.
- [15] J. Amin, et al., "An integrated design for classification and localization of diabetic foot ulcer based on CNN and YOLOv2-DFU models," *IEEE Access*, vol. 8, pp. 228586-228597, 2020.
- [16] M. Sharif, et al., "Recognition of different types of leukocytes using YOLOv2 and optimized bag-of-features," *IEEE Access*, vol. 8, pp. 167448-167459, 2020.
- [17] M. A. Khan, et al., "Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification," in 2019 Int. Conf. Comput. Inf. Sci. (ICCIS), IEEE, 2019.
- [18] M. A. Khan, et al., "Stomach deformities recognition using rank-based deep features selection," 2019, vol. 43, no. 12, pp. 1-15.
- [19] M. A. Khan, et al., "Multiclass Stomach Diseases Classification Using Deep Learning Features Optimization," 2021.
- [20] M. A. Khan, et al., "Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection," 2019, vol. 78, no. 19, pp. 27743-27770.
- [21] J. Naz, et al., "Recognizing gastrointestinal malignancies on WCE and CCE images by an ensemble of deep and handcrafted features with entropy and PCA based features optimization," 2021, pp. 1-26.
- [22] Z. Nayyar, et al., "Gastric tract disease recognition using optimized deep learning features," 2021, vol. 68, no. 2, pp. 2041-2056.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
- [24] Z. M. Lonseko, et al., "Gastrointestinal disease classification in endoscopic images using attention-guided convolutional neural networks," 2021, vol. 11, no. 23, p. 11136.

- [25] J. Amin, et al., "3D-semantic segmentation and classification of stomach infections using uncertainty aware deep neural networks," 2021, pp. 1-17.
- [26] I. M. Dheir and S. S. Abu-Naser, "Classification of Anomalies in Gastrointestinal Tract Using Deep Learning," *Int. J. Adv. Eng. Res.*, vol. 6, no. 3, 2022.
- [27] M. B. Haile, et al., "Detection and classification of gastrointestinal disease using convolutional neural network and SVM," 2022, vol. 9, no. 1, p. 2084878.
- [28] R. T. Sutton, et al., "Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images," 2022, vol. 12, no. 1, pp. 1-10.
- [29] W. Wang, et al., "Convolutional-capsule network for gastrointestinal endoscopy image classification," 2022.
- [30] S. Li, et al., "Effective high-to-low-level feature aggregation network for endoscopic image classification," 2022, pp. 1-9.
- [31] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4. Springer, 2006.
- [32] H. Borgli, et al., "HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Sci. Data*, vol. 7, no. 1, pp. 1-14, 2020.
- [33] J. T. Abraham, et al., "A deep-learning approach for identifying and classifying digestive diseases," *Symmetry*, vol. 15, no. 2, p. 379, 2023.
- [34] E. Sivari, et al., "A new approach for gastrointestinal tract findings detection and classification: deep learning-based hybrid stacking ensemble models," *Diagnostics*, vol. 13, no. 4, p. 720, 2023.
- [35] D. Mukhtorov, et al., "Endoscopic image classification based on explainable deep learning," *Sensors*, vol. 23, no. 6, p. 3176, 2023.



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

E-Active Workplace: An IoT based Solution to Improve Employee Workforce Efficiency Management

Hafiza Anisa Ahmed *, Syeda Aleeza Moazzam, Arisha Naz, Nabtahil Riaz, Mashal Noor

Department of Computer Science and Software Engineering, Jinnah University for Women, Karachi, Pakistan.
hafiza.anisa@juw.edu.pk, syedaaleeza519@gmail.com, arisha.ak01@gmail.com, nabtahilriaz@gmail.com,
mashalnoor215@gmail.com.

* Corresponding author: hafiza.anisa@juw.edu.pk

Article Citation:

Ahmed et al. (2024). "E-Active Workplace: An IoT based Solution to Improve Employee Workforce Efficiency Management". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract:

In this era, where the role of employees in an organization is universally recognized, there remains a significant gap in the practical implementation of strategies for effective employee management. ERM (Employee Resource Management) system is the best solution for this problem. Implementing an ERM system will help an organization manage employee's various duties on a single platform. Nowadays many companies have started to work on an ERM system to facilitate and engage employees for better results. IoT devices can play a vital role in monitoring and engaging employees. IoT integration within an employee management system revolutionizes workforce monitoring, by using sensors to detect employee movements, actions, and environmental data. Most of the organization are using IoT device from different perspective but not in a very helpful way. They are using IoT for environment monitoring, system monitoring, fitness trackers, and location trackers but using IoT for performance will help an organization to get a positive result. E- Active workplace is a solution with two different modules. Our first module will help the employees to request any on-desk resources without leaving their place. The fully automated leave management system contributes to creating an eco-friendly workplace. Appraisals are an important milestone in any employee's journey, providing useful insights, acknowledgment, and possibilities for progress but tracking employee full-year records is also difficult, our performance tracking feature will make it easier. Additionally, the system will monitor employee performance i.e., their motion, tone, and behavior by the use of IoT which will give them a better workspace where they can comfortably do their work, ask for any resource. Therefore, E-Active Workplace will provide a user-friendly platform created to simplify HR and workforce management tasks.

Keywords: Employee Resource Management, IoT, Employee performance, Alro Camera, Infrared Camera, Accelerometer/Gyroscope Sensor.

1. Introduction

Imagine a human without a backbone, similarly, you can imagine an organization without an employee [1]. Employees are an unseen force behind success and sustainability. Their hardworking skills and passion affect productivity, innovation, and customer satisfaction. Employees are the priceless resources that drive a company toward its objectives and guarantee its long-term success. Employee resource management (ERM) is a corporate management discipline that aims to increase the workforce's productivity and effectiveness. Managing employee performance, creating and putting into practice training plans, and overseeing compensation and benefits for employees are just a few of the many tasks involved [2].

An employee resource management system (ERM) is a digital technology designed to help in the effective management of an organization's employees. It has features like payroll, scheduling, employee databases, and performance monitoring [3]. Research has demonstrated that ERM systems streamline HR procedures, making them more precise and efficient. They facilitate better communication between HR teams and employees. However, setting up ERM systems can cause problems in firms. These include issues about data security, resistance to change from employees, and the need for proper training to use the system effectively. But to improve the productivity of this system we could use it

in a way whereby smart devices and sensors are incorporated into the office environment to improve employee productivity and well-being. Smart ID badges, temperature sensors, and occupancy trackers are examples of IoT-enabled devices that can improve workplace safety by monitoring employee well-being, guaranteeing adherence to safety procedures, and sending out instant notifications in the event of an emergency [4].



Published by
Foundation University
Islamabad.

Web: <https://fui.edu.pk/>

Many existing studies indicate that ERM systems, when integrated with IoT, can be highly beneficial for employee management and productivity. They use it in terms of environment monitoring, system monitoring, fitness trackers, location trackers, and employees' work, and experiences to collect real-time data [5]. But no one has worked for employee comfort or monitored employees for better use like employee performance by their motion, tone, and behavior. As employee performance is the key element of an organization productivity, our system will use it for better interaction and to make employees engaged.

An E-Active Workplace aims to provide a platform to cooperate with employees and human resources. This includes activities such as managing employee resources, monitoring daily tasks, tracking employee performance, managing leaves, and maintaining employee information. Our goal is to develop a comprehensive platform where employees can easily submit requests for their desk-related issues, ensuring prompt response. This system will automate task management, making it easier for employees to handle their daily tasks and making it possible for managers to effectively track task completion. Performance reviews will be expedited at assessment time by utilizing lead remarks and task submission times. To track performance in real-time, we are also integrating IoT. A paperless leave management system will simplify leave requests and provide fast responses accepted, refused, or pending as well as employee profiles that compile essential information. IoT will add uniqueness to our system, focusing on different areas for employee resources. We will use an accelerometer sensor and a gyroscope sensor to check the movement of the employee. If the employee gets distracted, frustrated, or tense because of the need for any resource, it will detect that. For employee task performance, we will use a camera for better monitoring of employee performance. Bluetooth and a microphone will also be used for employee performance by recording the employee's voice tone. For performance, we will use an Alro camera and IR (Infrared Camera) for face-to-face interaction and for checking conversations and recording movement. All sensors mainly work on employee performance, as it is our key element.

The main contributions of this research can be summarized as follows:

- a) E-Active Workplace streamlines HR processes, automating tasks like leave management, employee on-desk resources, monitoring daily tasks, and employee profile access.
- b) The system also simplifies performance tracking of employees using IoT devices for real-time insights that will help enhance efficiency and data accessibility.

2. Literature Review

Murthy, M. Narayana, and P. AjaySaiKiran [6] proposed an office automation system based on the Raspberry Pi model B and IoT technology that uses an Android mobile app to control room temperature, humidity, and electrical loads. They highlight the shortcomings of conventional systems while emphasizing the advantages of Wi-Fi-based automation. The architecture of the system includes a Raspberry Pi, sensors, an Android app, and electrical appliances. It monitors environmental conditions with sensors and activates loads accordingly. Remote control and monitoring are possible using the Android app. Deploying Raspbian OS, attaching sensors and loads, and automating with a lookup table technique are all part of the implementation. The system offers improved comfort and energy efficiency in offices and can be controlled from anywhere within the network.

Dr. A. Narasima Venkatesh [7] explores the Internet of Things (IoT) and its impact on the workplace and human resource management. The paper discusses how the Internet of Things is transforming the way businesses operate, particularly in terms of human resource management. It emphasizes the shift toward mobile and flexible work arrangements for the future generation of employees named "GenMobile." The increase in the number of Internet of Things (IoT) devices such as mobile phones, fitness trackers, and tracking devices for location is changing how workers work and how their experiences are monitored. The paper also underlines the importance for organizations updating their physical environments and technological design to facilitate IoT ecosystems. It cites examples of companies such as Intel embracing digital workspace to increase productivity and lower real estate costs. Overall, the author suggests that IoT will grow and develop into an integral component of the workplace, impacting many HR functions.

Prof. P. R. Rodge et al. [8] developed a low-cost office automation system that uses IoT and Raspberry Pi to control electrical devices in an office. A visitor counter is also included in the system to help with resource management. IoT enables remote monitoring and control via a smartphone app, offering consumers with a user-friendly interface. The system's architecture involves relay circuits connecting devices to the Raspberry Pi, and a mobile app interacting with the Raspberry Pi via the cloud. The mobile application includes features such as turning appliances on and off and tracking visitor counts. Cloud connectivity is provided through Microsoft Azure cloud services. The results show that the system is a simple and effective approach to control workplace equipment, minimize electricity consumption, and track visitor counts.

Xun Wu [9] studies the integration of Internet of Things (IoT) technologies into systems for human resources management. The study highlights the limitations of traditional human resource management, emphasizing the use of technological solutions to improve accuracy and efficiency, particularly in the context of telecommuting and post-epidemic workplace changes. The paper discusses how IoT may help with planning, talent recruitment, training,

performance management, compensation calculation, employee interactions, and career development planning. However, the study recognizes the significant expenses of IoT deployment and argues that high-tech firms and those requiring considerable data collecting are best suited for it.

Bhairavi Gyandev Kapadnis et al. [10] proposed an Employee Management System (EMS) as a comprehensive solution to address human resource challenges in organizations. The system aims to fulfill a wide range of managerial requirements by providing user-friendly features for strategic planning and ensuring proper human resource allocation for future objectives. The EMS also features remote access capabilities, allowing busy executives to manage their personnel while on the go, enhancing resource management. This system's sophisticated performance tracking and analysis capability, including employee sales verification, is a notable advantage. It also includes an effective time tracking mechanism, which saves the organization time and money. The system's algorithm streamlines employee performance monitoring, tracking, and analysis through a user-friendly interface. To ensure accuracy and transparency, administrators can manage personnel information, establish regions, and review sales reports. Furthermore, the proposed EMS is an appealing choice for businesses seeking better human resource management. Its successful implementation is projected to improve overall employee performance and resource utilization by streamlining management tasks, simplifying record-keeping, boosting transparency in sales analysis, and streamlining managing processes.

There are many existing employee resource management systems that are using IoT technology but they are focusing on basic monitoring such as attendance tracking, computer sleep timers and detecting whether the employee is sitting in his seat or not. These systems are not working on deep insights into employee behavior or well-being. Compared to the existing system, the proposed system works on detailed information and collects comprehensive data such as employees' physical movements, their workplace interactions, and emotional tone to reveal employees' work and health. This can not only improve the performance of the employees but also increase the productivity towards the organization.

3. Proposed System

Many companies struggle with effectively managing their employees' time and resources. This often leads to inefficient use of employee time, missed deadlines, and decreased productivity. Inefficient employee resource management can result in increased costs and lower profitability, especially for businesses that rely on billable hours. Additionally, employees may feel stressed and overworked, leading to a higher rate of turnover and decreased employee satisfaction.

An E-Active Workplace seeks to provide a platform for employees and human resources to collaborate. This comprises duties such as managing employee's resources, monitoring daily tasks, tracking employee performance, managing leaves, and preserving employee information. In Figure 1, there is a look at the dashboard which shows all of our features. The following features are discussed in detail below.

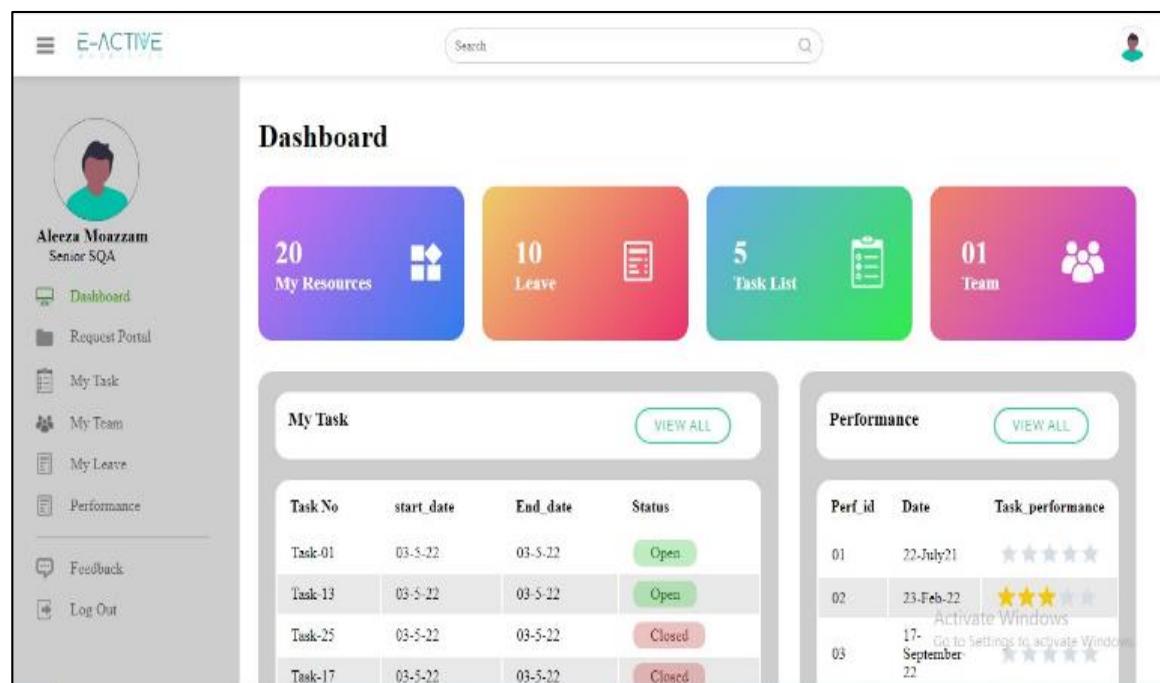


Figure 1: Employee Dashboard

1.1. Resource Management

To cater to employees' on-desk needs, we want to make this system a helpful platform from which employees can generate requests for any needs they face while working. An employee can simply log in to the system and request the request portal as shown in Figure 2, and that request will be resolved in the shortest possible time. For an employee who is working and during his work, he faces a need for printer paper but can leave the system because of urgent work in this situation, he can simply use the resource request module of E-Active Workplace. The employee will request any desired item through our proposed system and that item will be sent to the employee's desk. As shown in Figure 3, there are resources that the employee is currently using. This way, the employee will not need to do paperwork to request the resource or call the HR department to request the resource.

The screenshot shows the 'Request Portal' section of the E-Active Workplace system. On the left, a sidebar displays the user profile of 'Aleeza Moazzam' (Senior SQA) and a navigation menu with options: Dashboard, Request Portal (which is selected), My Task, My Team, My Leave, Performance, Feedback, and Log Out. The main content area is titled 'Request Portal' and contains four dropdown fields: 'Requested Item *', 'Building *', 'Floor *', and 'Room *'. A message at the bottom right says 'Activate Windows Go to Settings to activate Windows.'

Figure 2. Employee Request Portal

The screenshot shows the 'My Resources' section of the E-Active Workplace system. The sidebar is identical to Figure 2. The main content area is titled 'My Resources' and features a table with columns: Resource ID, Employee ID, Resource Name, Type, Date, and Description. All rows in the table show 'Resource ID': 101, 'Employee ID': emp_123, 'Resource Name': Drive, 'Type': Tool, 'Date': 06-7-22, and 'Description': -----. A message at the bottom right says 'Activate Windows Go to Settings to activate Windows.'

<input type="checkbox"/>	Resource ID	Employee ID	Resource Name	Type	Date	Description
<input type="checkbox"/>	101	emp_123	Drive	Tool	06-7-22	-----
<input type="checkbox"/>	101	emp_123	Drive	Tool	06-7-22	-----
<input type="checkbox"/>	101	emp_123	Drive	Tool	06-7-22	-----
<input type="checkbox"/>	101	emp_123	Drive	Tool	06-7-22	-----
<input type="checkbox"/>	101	emp_123	Drive	Tool	06-7-22	-----

Figure 3. Resource Management

1.2. Task Management

Tracking employees' tasks daily is a tedious and difficult task to remember. Monitoring employee tasks on an automated system will help an employee easily recall their task as shown in Figure 4, and the lead can also have a look at each employee's task list and status using this system.

Figure 4. Task Management

1.3. Performance Management

In every organization, appraisals have to be done at the end of the year, but recalling each employee's performance on each task is difficult. In an E-Active Workplace, performance can be easily measured as shown in Figure 5, based on the time of task submission and by the remarks of the lead after reviewing the task.

Figure 5. Performance Management

1.4. Employee Information

Employees' information is only mentioned on their CVs, and that CV can't be looked at every time of need. We want to display the employee's profile on this system as shown in Figure 6, which will display every important detail related to the employee. For employee skills, if a team lead wants to apply for a task focusing on a particular skill, our proposed system will display the employee profile with their skills and their performance on previous projects.

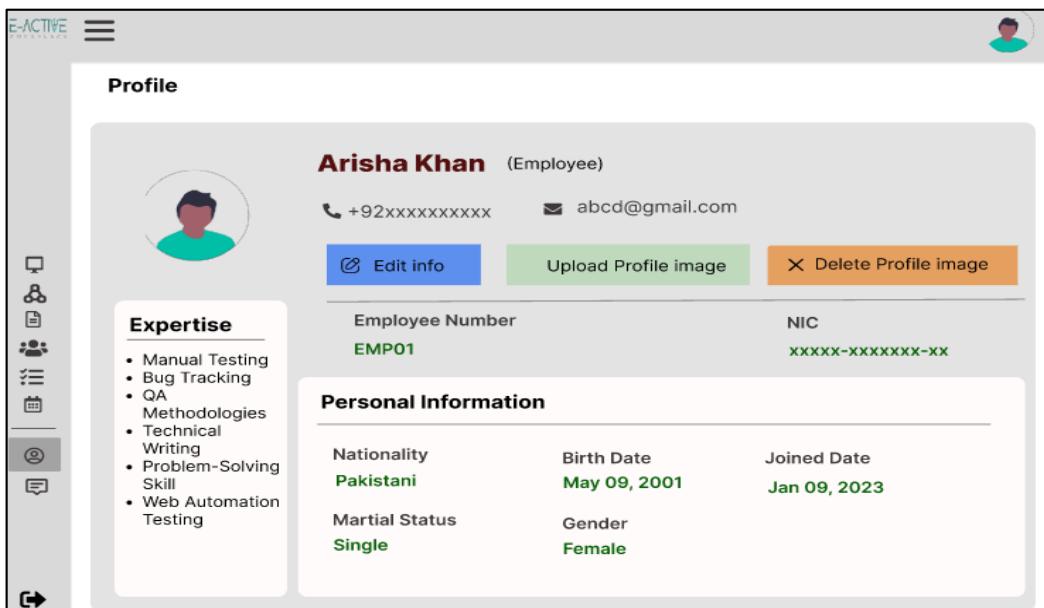


Figure 6. Employee Information

1.5. Leave Management

For reducing paperwork and delays in leave replies, the leave management feature is the best solution. An employee can apply for a leave application using this system as shown in Figure 7, and submit it to the desired person, after which the employee's leave notification will be received by the team lead. The lead can decide whether the leave is accepted, rejected, or pending.

Leave ID	Reason	Apply Date	From Date	To Date	Type	Description	Status
101	Vacation Leave	06/02/2023	06/02/2023	06/02/2023	Full Day	Do to some my reason	Approved
101	Marriage Leave	06/02/2023	06/02/2023	06/02/2023	Full Day	Do to some my reason	Pending
101	Casual Leave	06/02/2023	06/02/2023	06/02/2023	Full day	Do to some my reason	Activate Windows Go to Settings to activate Windows.

Figure 7. Leave Management

4. Integrating E-Active Workplace System with IoT

By integrating the E-Active Workplace System and IoT technologies, we are creating a smarter and more efficient work environment. This means that we can use sensors and cameras to watch how employees walk, talk, and interact, allowing them to request resources, manage tasks, and improve their performance more easily.

4.1 Accelerometer /Gyroscope Sensors

Accelerometer and gyroscopes are found in almost all current mobile phones as well as many robotic applications. They may help us in determining how rapidly and in which direction our equipment is moving, as well as how quickly it is

turning. Accelerometer measure devices' linear acceleration, or acceleration along an axis. Gyroscopes, on the other hand, detect angular velocity, or how rapidly the body is turning, using the Coriolis Effect rather than acceleration [11].

4.2 Arlo Smart Camera

Arlo Smart provides intelligent services that give you more control over how you monitor, manage, connect, and respond to events in your home or company [12]. With Arlo Smart, you can improve your security experience by receiving more intelligent motion alerts, allowing you to take immediate action, sounding an alarm, or calling emergency services.

4.3 Bluetooth Microphone

Bluetooth microphones provide long-distance recording at a minimal cost. They are portable and suited for on-location use when capturing sound or video content because they provide wireless online connections [13].

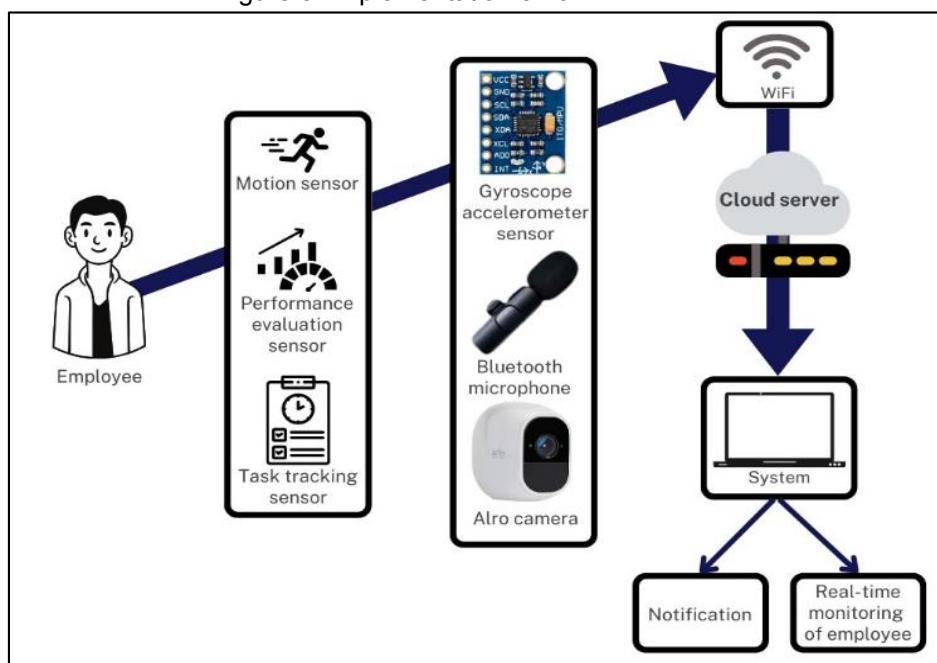
4.4 Infrared Sensor

Infrared (IR) sensors are most commonly used in wireless technology, which includes remote control functions and detection of surrounding objects or obstacles [14]. IR sensors are now widely used in motion detectors, which are used in building services to turn on lamps or in alarm systems to detect unwelcome guests [15].

5. Implementation

The IoT integrating sensors and technologies are designed to improve and maximize employee performance, which is the fundamental goal of our system as illustrated in Figure. 8. We can get important information into employee behavior, productivity, and well-being by collecting data from these sensors. This data can be processed and analyzed using modern algorithms and machine learning approaches to find patterns and trends.

Figure 8. Implementation of IoT



Accelerometers and gyroscopes are sensors that detect physical movements and orientation. In this system, accelerometers measure linear acceleration, which can be used to identify actions like walking or typing, and gyroscopes measure angular velocity and orientation changes, which can be used to identify tilting or shaking. These sensors' combined data enables machine learning algorithms to recognize specific behaviors. By using decision trees or Recurrent Neural Networks (RNNs), the system can distinguish between typing at a desk and physical exercise in the workplace, enhancing employee safety and well-being. Time series data, such as accelerometer and gyroscope readings, can be effectively analyzed using Long Short-Term Memory (LSTM) to identify movements and trends over time.

Infrared sensors detect heat and motion, enabling the system to monitor employee presence in specific areas. When integrated with machine learning, these sensors may determine occupancy patterns, allowing for more efficient resource allocation. Clustering algorithms such as K-Means can group employees based on their sensor data, patterns of

behavior, workspace utilization, and assisting in office space management. It is useful for locating various work styles or activity clusters.

Bluetooth microphones capture audio data, and machine learning models can be used to monitor employee interactions and ensure compliance with workplace policies. Sentiment analysis algorithms, in conjunction with Bluetooth data, may determine the emotional tone of interactions. Natural language processing (NLP) models can assess whether or not conversations in meeting rooms are constructive or if conflicts arise, contributing to a more peaceful work environment.

Since our proposed system is working on employee data, data privacy and security are key concerns. For this, data encryption algorithms such as AES-256 data privacy and TLS will be integrated into the system for secure data transmission.

A centralized system must be used to gather, process, and analyze the data from all of these sensors by using various algorithms. Our IoT-enabled solution is designed to give an integrated approach to employee performance management by utilizing multiple sensors, devices and algorithms to monitor, analyze, and optimize various aspects of employee's resources, resulting in a more productive and efficient workplace.

6. Challenges

There are potential challenges with our proposed IoT-based system. Employees of the organization will feel uneasy when they are monitored more and their actions are constantly monitored. Also, they may be suspicious of their personal data and details being misused by others or the system's automated insights may provide an unfair assessment of performance.

For these challenges, employers will be provided with detailed information on what data is being collected, and how it will be used, and open channels for feedback will be ensured. In addition, offering employees the option to voluntarily participate or adjust the level of monitoring will allow choice to build trust and reduce resistance.

7. Conclusions

In the ever-evolving world of workplace management, our system emerges as a pioneering solution that seamlessly integrates IoT technology to revolutionize employee management. Imagine a workplace where employees can easily request resources, track tasks, and receive real-time performance feedback via an easy-to-use platform. With sensors and cameras tracking their motions, tone, and interactions, we not only increase productivity but also create a supportive environment in which employees thrive. This isn't just an idea; it's a system about how innovation and technology can help organizations reach new heights while ensuring their most important assets which are employees, to feel valued and supported every step of the way.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

8. References

- [1] Nanayakkara, Sanuji, et al. "A Web Based Employee Management System." International Journal of Engineering and Management Research 12.5 (2022): 82-89.
- [2] Adamu, Abubakar. "Employee Leave Management System." FUDMA Journal of Sciences 4.2 (2020): 86-91.
- [3] Loh, Chun Kang, and Rozanawati Darman. "The Implementation of Employee Management System on SMC Technology Sdn. Bhd." Applied Information Technology And Computer Science 3.2 (2022): 289-306.
- [4] Nalini, A. "Employee Task Allocation System." International Journal of Engineering and Management Research 9.2 (2019): 225-228.
- [5] Das, Mr Surajit, and Ranajit Senko. "Online Employee Management System." (2021).
- [6] Murthy, M. Narayana, and P. AjaySaiKiran. "A smart office automation system using raspberry pi (model-b)." 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018.
- [7] Venkatesh, Dr A. Narasima. "Connecting the dots: Internet of Things and human resource management." American International Journal of Research in Humanities, Arts and Social Sciences, ISSN (Print) (2017): 2328-3734.
- [8] Rodge, P. R., et al. "IoT based smart interactive office automation." International Research Journal of Engineering and Technology (IRJET) 4.04 (2017): 982-986.
- [9] Wu, Xun. "Research on human resource management system based on Internet of Things technology." International Journal of Frontiers in Sociology 3.10 (2021): 39-42.

- [10] Kapadnis, Bhairavi Gyandev, Aishwarya Sunil Ghumare, and Kaveri Bhausaheb Deore. "Analysis, Performance Monitoring and Tracking System for Employees."
- [11] Ahmed Bhuiyan, Rasel, et al. "A robust feature extraction model for human activity characterization using 3-axis accelerometer and gyroscope data." Sensors 20.23 (2020): 6990.
- [12] Beugin, Yohan. "Building a Secure and Privacy-Preserving Smart Camera System." (2021).
- [13] Chunhakam, Puripak, Phakakorn Panpho, and Paramote Wardkein. "Bluetooth Breathing Sound Detection Device Based on Time and Frequency Domain Analyze." Proceedings of the 12th International Conference on Biomedical Engineering and Technology. 2022.
- [14] Malkhede, Payal, Mayuri Chawala, and Gaurav Sambhe. "Employee/Student Attendance Temperature Monitoring Using RFID Cards." Annals of the Romanian Society for Cell Biology (2020): 360-367.
- [15] Holten Møller, Naja, et al. "Can workplace tracking ever empower? Collective sensemaking for the responsible use of sensor data at work." Proceedings of the ACM on human-computer interaction 5.GROUP (2021): 1-21.

Human Obesity and Coronavirus: A Bibliometrics Review of Emerging Literature

Mubbasher Munir ^a Ganesh Kumar ^{b,*} Zahra Tul Amani Zakaria ^c Atif Amin Baig ^d

^a Department of Business Analytics, Sunway Business School, Sunway University Malaysia.
(mubbasherm@sunway.edu.my).

^b School of Engineering and Technology, Sunway University, Malaysia
(ganeshk@sunway.edu.my).

^c Faculty of Informatics and Computing, Universiti of Sultan Zainal Abidin, Terengganu, Malaysia

^d International Medical School, Management and Science University, Malaysia.

* Corresponding author: ganeshk@sunway.edu.my

Abstract:

Obesity is a serious threat worldwide, with global eating disorders and coronavirus pandemic in the coming era. This study is developed through web of sciences (WoS) databases of scientific research publications using the PRISMA framework. Finally, 477 research articles were used to analyze the following literature for best co-occurrences, clustering and networking. The results found that most developed nations are working with developed countries but sometimes are connected with developing countries. Same looks with author's and institution's behaviors. Recent literature has highlighted significant connectivity among countries, institutes, documents, and key authors regarding all possible bibliometric analysis tools. Global literature clearly shows how globalization works with strong research connections and established research connectivity.

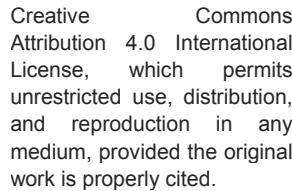
Keywords: Coronavirus, COVID-19, Co-occurrence, SARS-COV-2, Infection, Disease, Health, Outbreak, Obesity

1. Introduction

In December 2019, cases of pneumonia of unknown cause were seen in Wuhan, China [1]. It was discovered to be caused by a previously unknown beta coronavirus named 2019-nCoV [2]. Later on, 2019-nCov was renamed as SARS-CoV-2 [3]. Most of the early confirmed cases were linked to the Huanan Seafood Wholesale Market [4]. On January 30, 2020, WHO declared the novel coronavirus outbreak a Public Health Emergency of International Concern (PHEIC). On March 11, 2020, WHO declared Covid 19 a pandemic [5]. As of January 21, 2021, there are 95,321,880 confirmed cases and 2,058,227 deaths worldwide [6].

The most common Covid 19 symptoms include fever and cough. Other common symptoms include fatigue, myalgia, expectoration, anorexia, chest tightness, dyspnea, nausea and vomiting, diarrhea, headache [7]. Additionally, commonly observed complications in [8] are sepsis, respiratory failure, ARDS, heart failure, and septic shock. [8] states that the patients with diabetes or coronary heart disease had a higher probability of death in the hospital due to Covid 19. There is a link between obesity and cardiovascular diseases, and diabetes [9]. Besides, especially in patients with BMI > 35, obesity has been shown to affect disease severity [10].

Covid 19 is a disease caused by the novel coronavirus, later named Sars-cov-2 [3]. Early patients were linked to Huanan Seafood Wholesale Market in Wuhan, China [4]. Covid 19, where the first cases were seen in China, spread rapidly worldwide [6]. On January 30, 2020, WHO stated that the novel coronavirus epidemic was qualified as PHEIC, and on March 11, 2020, declared Covid 19 as a pandemic [5]. There were 95,321,880 confirmed cases and 2,058,227 deaths worldwide on January 21, 2021 [6].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

Fever and cough were frequently observed in patients with Covid 19. Besides fever and cough, symptoms such as fatigue, myalgia, expectoration, dyspnea, nausea and vomiting, diarrhea, headache can be seen [7]. Covid 19 can lead to conditions such as sepsis, respiratory failure, ARDS, heart failure, and septic shock in patients [8]. In a study with 1591 ICU patients, it has been shown that the ICU is needed for most patients is acute hypoxic respiratory failure [11].

Another study with 1099 patients shows that ground-glass opacity is the most common radiological finding [12]. Usually, Sars-Cov-2 is transported to the organism by droplets entering the respiratory tract [13,14]. WHO [15] defines obesity as "abnormal or excessive fat accumulation that may impair health", where usually body mass index (BMI) is used to detect. WHO [15] classifies Obesity if BMI is higher than 30. In 2016, globally, obesity was present in 13% of adults [15].

Overweight or obese patients, especially patients with BMI > 35, have a higher chance of increased Covid-19 disease severity [10]. It is shown that the severity of the disease increased as the BMI. Patients with obesity have a higher risk of hypertension, cardiovascular disease, type 2 diabetes [16]. It is stated that obesity affects the immune system and disrupts the inflammatory-related immune cells. With the problems with inflammation, patient conditions become more severe as Covid-19 related inflammation increases [17].

1.1. Literature Review

Bibliometric analysis can help us review the literature in a particular area. For instance, in a recent study on the Environmental Kuznets Curve, an analysis was made using keywords, citations, journal, and country information. 2384 records collected from the Web of Science database between 1999 and 2019 were used. As a result of the study, inferences were made about what should be considered in future studies [18].

After Covid 19 became an important part of our lives, a lot of research has been done. The analysis of the literature continues throughout the pandemic. Some of the studies conducted, such as [19,20], Covid 19 literature has been evaluated more generally. Some studies include [21–24], COVID-19 and specific topics have been analyzed.

In the study by [19], 1809 records between December 2019 and March 2020 were retrieved from the WHO database and PubMed with "novel coronavirus 2019", "coronavirus 2019", "COVID 2019", and "COVID 19" keywords. After removing duplicates and applying other filters, 564 records were used. The data obtained were especially evaluated according to their countries. Articles published between 2003 and 2020 were collected from Web Of Science Core Collection: Citation Indexes. In the data retrieval process, "coronavirus or Middle-East-Respiratory-Syndrome or Severe-Acute-Respiratory-Syndrome or 2019-nCoV or COVID-19 or SARS-CoV-2" filter was used. Additional filters, such as document type and index, are applied. The records obtained were analyzed by the trend, category, authors, organizations, journals, and keywords.

Records collected from Scopus, BioRxiv, and MedRxiv in June 2020 were used [21]. Besides similar bibliometric analysis, this study also has short reviews on publications. 276 articles collected in June 2020 were used, and the bibliometric analysis was made with author, keyword, and country information. The articles were reviewed briefly, with the analysis made. In [22], COVID-19 research in business and management has been reviewed. The studies were reviewed under clusters determined with the help of bibliometric analysis. 12068 papers retrieved from Scopus and WoS databases. As a result of applying additional filters, 107 papers were used. With the help of a co-word analysis with 107 articles, 4 themes and 18 sub-themes were determined. In another study, a bibliometric analysis of research on mental health at the time of the H1N1, Ebola, and Covid 19 outbreaks was performed [23].

3524 records were retrieved from Embase, PubMed, and Scopus databases on related keywords from the beginning of each outbreak until August 2020. For each disease, data such as number of documents, number of open-access documents, number of documents by country, the proportion of documents in high impact journals, proportion of international collaboration documents were extracted. Publications at times of the outbreak were compared with the data obtained [23].

In the [25], publications about Covid 19 and Obesity were evaluated. Instead of bibliometric analysis, systematic review and meta-analysis approaches were used in this study. MEDLINE, EMBASE, Scopus, Web of Science, CENTRAL, OpenGrey, medRxiv, and bioRxiv databases were used and accessed in May 2020. In the search, Covid 19 related keywords such as "coronavirus", "SARS-COV-2", and Obesity related keywords such as "obesity" and "hyperphagia" were used. "Clinical studies (1) that reported obesity prevalence or outcomes, and (2) that were performed on COVID-19 patients" were used as the main criteria. PRISMA framework used for the article selection. From the initially retrieved 1493 articles, only 61 were revied in the study. From the 61 studies, 270241 patients meeting the criteria were used. The studies were evaluated under four topics, "ICU admission and critical illness", "Severe disease", "Mortality", and "Positive test". Protocol for the study was registered on PROSPERO (CRD42020184953).

A systematic review was conducted in another study on Covid 19 and Obesity. In this study, "longitudinal cohort studies or randomized controlled trials which observed the effect of BMI to the requirement of in-hospital critical care in patients infected with 2019- novel coronavirus (COVID-19)" used as inclusion criteria. Cochrane, MEDLINE, EMBASE, and PubMed databases were used and accessed in April 2020. Covid 19 and obesity related keywords such as "covid-19", "morbidity", "sars-cov-2" used in the initial database search. Initially, 573 articles were retrieved. PRISMA framework used for data screening. After the data exclusion, only 3 articles were reviewed. Protocol for the study was registered on PROSPERO database (CRD42020183068) [26].

In another systematic review and meta-analysis study on Covid 19 and Obesity, particular focus is placed on the effect of BMI on the ICU admission rate. Initially, 15168 articles were retrieved from Medline, Embase, CENTRAL, Scopus, and Web of Science. Keywords such as "covid 19", "coronavirus", "SARS-COV-2" are used in the retrieval process. PRISMA framework used in the screening process, and only 24 studies were reviewed. "authors, publication year, digital object identifier, study site, study design, protocol number, age, gender distribution, number of patients in each reported BMI range, number of patients with IVM in each reported BMI range, number of patients with ICU admission in each reported BMI range, odds ratios for IVM and ICU admission regarding BMI groups" data has been extracted from 24 studies and reviewed according to. The review protocol was registered on PROSPERO database (CRD42020185980) [49].

1.2. Research Objectives

The objectives of this research are to reach the scientific literature on the topic of coronavirus and obesity using bibliometric analysis through multivariate data analysis, clustering and modern graphical representation methods and to come up with a set of parameters that can support researchers and practitioners.

2. Methodology

2.1. Study Design and Data Descriptions

This study aimed to analyze coronavirus published documents using bibliometric analysis to investigate the effects on global health. The most recent coronavirus epidemic introduced uncertainty and severe risk to humanity, but at the same time, significantly reshaped society overall. The study uses a scientific production approach to examine previous literature [27].

The current study has been designed as a result of a review of the research literature on Obesity and COVID19, followed by a comparative survey of works in the literature that were structured into research axes to look for interfaces. In this scope, bibliometrics, a method for evaluating academic production, can be thought of as systematic, intuitive, and reproducible, allowing for easier understanding of scientific and technical data cite this article [28]. In terms of methodology, the techniques employed in this analysis are close to those used [28,29]. But, the study scope was limited to published articles and a particular time frame, from 2019 to 2021.

Bibliometrics is a collection of quantitative tools to evaluate science and technical literature [30,31]. Most scholars commonly accept that bibliometrics owes its formal creation primarily to the pioneers of Price DJD and [32] and extend [33], But Alan Pritchard first used the word "bibliometrics" in his article, Statistical Bibliography or Bibliometrics? Published in the 1969 Journal of Documentation [34], which described books and other media of communication as the application of mathematics and statistical methods.

Through measuring objects from physical publication units, bibliographic cites, etc., Bibliometrics delineates the study entity [35]. The common functional application of bibliometric methods in measuring study output is citation analysis and co-citation analysis. There has been considerable discussion about the importance of citation analysis to assess the effect of the study because of the difficult citing actions. However, there is ample proof that the comparison motivations are not so distinct or "randomly given" to the degree that the citation phenomenon loses its function as a valid measure of effect. In certain cases, citation analysis can provide a clear empirical success measure when extended to the whole job [36].

Co-citation review is the frequency with which the later literature cites two objects from previous literature [37,38]. This illustrates the association between the documents. The more co-citations two documents get, the greater their intensity of co-citation, and the more likely they are to be semantically connected [38].

2.2. Data Collection and Sources

Thomson Reuters' ISI Web of Science (WoS) is considered a significant source of data for bibliometric research in the sciences [30,39]. Its records are more reliable and systematic than other databases such as WoS [40,41], enabling us to retrieve title words and author titles, most important, referenced references for our bibliometric analysis. Therefore, we choose WoS as the data source for our research in this paper. Results from COVID 19 research are not presented only in scientific journals [42]; many international conferences, especially during the formative years.

In recent years, a large amount of research into Coronavirus has been published [43–47]. We followed the Preferred Reporting Items for (PRISMA) approach for bibliometric analysis of covid-19 and Obesity [27,45]. We have collected the data for this study from the Web of Science Core Collection database. On 19.01.2021 at 4:10 PM (GMT+3), when the data was retrieved, there were 956 publications published in 2019 - 2021, including the keyword obesity from a total of 93393 articles written about covid 19, coronavirus, SARS-COV-2, or ncov19. Our reason for restricting the years is to exclude the research related to the older coronavirus derivatives before 2019. Other than that, no restrictions are used for search. The following keywords have been used for this study: "COVID 19" OR "coronavirus" OR "SARS COV 2" OR "NCOV19" AND "OBESITY" AND publication years" 2021" OR "2020" OR "2019". Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A & HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI.

Considering the seriousness of the situation, it can be seen that Covid 19 and related subjects were the focus of scientific studies during this period. As explained in detail in the dataset section, initially, 93393 records were retrieved from the Web of Science Core Database related to covid-19. As a result of the assessments made from these records, 477 were used in this study. This bibliometric analysis was carried out to examine Covid 19 and Obesity studies according to country, journal, topic, and keyword aspects.

Co-citation analysis is a basic type of citation that can offer a usefully detailed image of the speciality's intellectual essence, the pace and course of its evolution, and the number and name of its main individuals [30]. The work of [48] suggested that In comparing the precision of various search methods on the scale of millions of papers, the co-citation analysis tool performs better than the direct citation analysis approach. The concept of co-citation analysis that describes the nature of intellectual information in terms of networks of co-cited references is the basis of several science analysis tools [38]. Co-citation review is one of the most common and powerful methods for identifying key papers in a body of literature [49]. Some of these points will have to be more established as we move along. A data visualization technique is used in the work of this paper to investigate the structure and complexities of the co-citation network.

Two forms of co-citation analysis exist the author's co-citation analysis and the keywords' co-citation analysis. The method of paper co-citation makes a reasonably accurate assertion as to how wide-ranging the records of an author are [50]. This study focuses on document co-citation analysis, with a primary goal of identifying the intellectual structure of coronavirus research documents in terms of the groupings formed by accumulated co-citation trails in scientific literature.

After obtaining the data, we filtered it according to the document type. At this stage, we removed the Correction, Editorial Material, Editorial Material; Early Access, Letter, Letter; Early Access, Meeting Abstract, News Item, Review, Review; Early Access, and Article; Early Access type documents from the data set. As a result, a total of 477 studies remained in Article type.

The following analyses were performed using the source, country, affiliation, citation, and keywords of 477 research publications obtained due to the process described above. In the bibliometric data obtained from WoS, the month and year information of the publications are found separately. For the periodic analysis part of the study, the date of the publications was obtained by combining this month and year information. 102 documents that do not specify publication month are not included in the periodic analysis section.

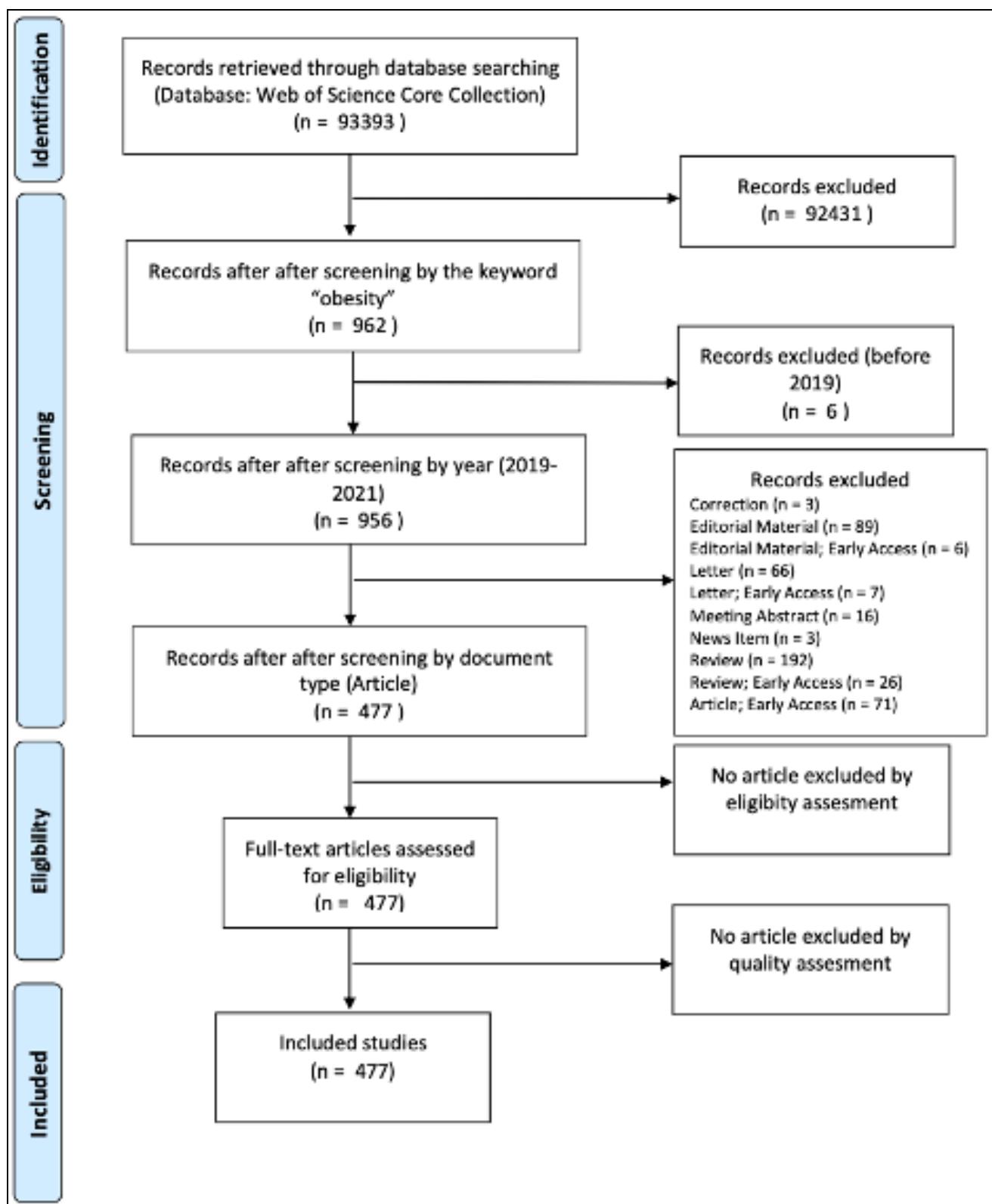


Figure 1. PRISMA framework

3. Results

3.1. Description of used documents

Table 1 shows the article's main information, which can represent a summary of available literature till 19.01.2021 at 4:10 PM (GMT+3). The results are based on 477 research published articles from 282 sources. On average, these documents have found 12.56 citations per article from 14,037 references, including 1091 author's keywords with 6.263 on average citations per year per document.

Table 1. Description of used documents

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2019:2021
Sources (Journals, Books, etc.)	282
Documents	477
Average years from publication	0.964
Average citations per document	12.56
Average citations per year per doc	6.263
References	14037
DOCUMENT TYPES	
Article	477
DOCUMENT CONTENTS	
Keywords Plus (ID)	749
Author's Keywords (DE)	1091
AUTHORS	
Authors	4326
Author Appearances	4518
Authors of single-authored documents	25
Authors of multi-authored documents	4301
AUTHORS COLLABORATION	
Single-authored documents	25
Documents per Author	0.11
Authors per Document	9.07
Co-Authors per Documents	9.47
Collaboration Index	9.52

Table 2 and figure 2 represent the facts that "obesity" research journal is secured the highest position which is 5.45 % in obesity and covid19 related research articles among all the other research journals, in second position research journal name "nutrients" which have been produced 3.56 % research articles during 2019 to January 2021 and Plos one secured the third position to produce research articles and so on.

3.2 Most Relevant Sources

Table 2. Most Relevant Journals (Sources)

Sr.#	Journals (Sources)	Articles	%
1	OBESITY	26	5.450734
2	NUTRIENTS	17	3.563941
3	PLOS ONE	15	3.144654
4	CUREUS	10	2.096436
5	DIABETES, METABOLIC SYNDROME CLINICAL RESEARCH & REVIEWS	9	1.886792
6	MEDICAL HYPOTHESES	9	1.886792
7	INTERNATIONAL JOURNAL OF OBESITY	6	1.257862
8	JOURNAL OF CLINICAL MEDICINE	6	1.257862
9	METABOLISM-CLINICAL AND EXPERIMENTAL	6	1.257862
10	CLINICAL OBESITY	5	1.048218
11	DIABETES CARE	5	1.048218
12	OBESITY RESEARCH \& CLINICAL PRACTICE	5	1.048218
13	DIABETES RESEARCH AND CLINICAL PRACTICE	4	0.838574
14	FRONTIERS IN PUBLIC HEALTH	4	0.838574
15	INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	4	0.838574
16	NUTRITION METABOLISM AND CARDIOVASCULAR DISEASES	4	0.838574
17	OBESITY REVIEWS	4	0.838574
18	OBESITY SURGERY	4	0.838574
19	BMC PUBLIC HEALTH	3	0.628931
20	BMJ-BRITISH MEDICAL JOURNAL	3	0.628931

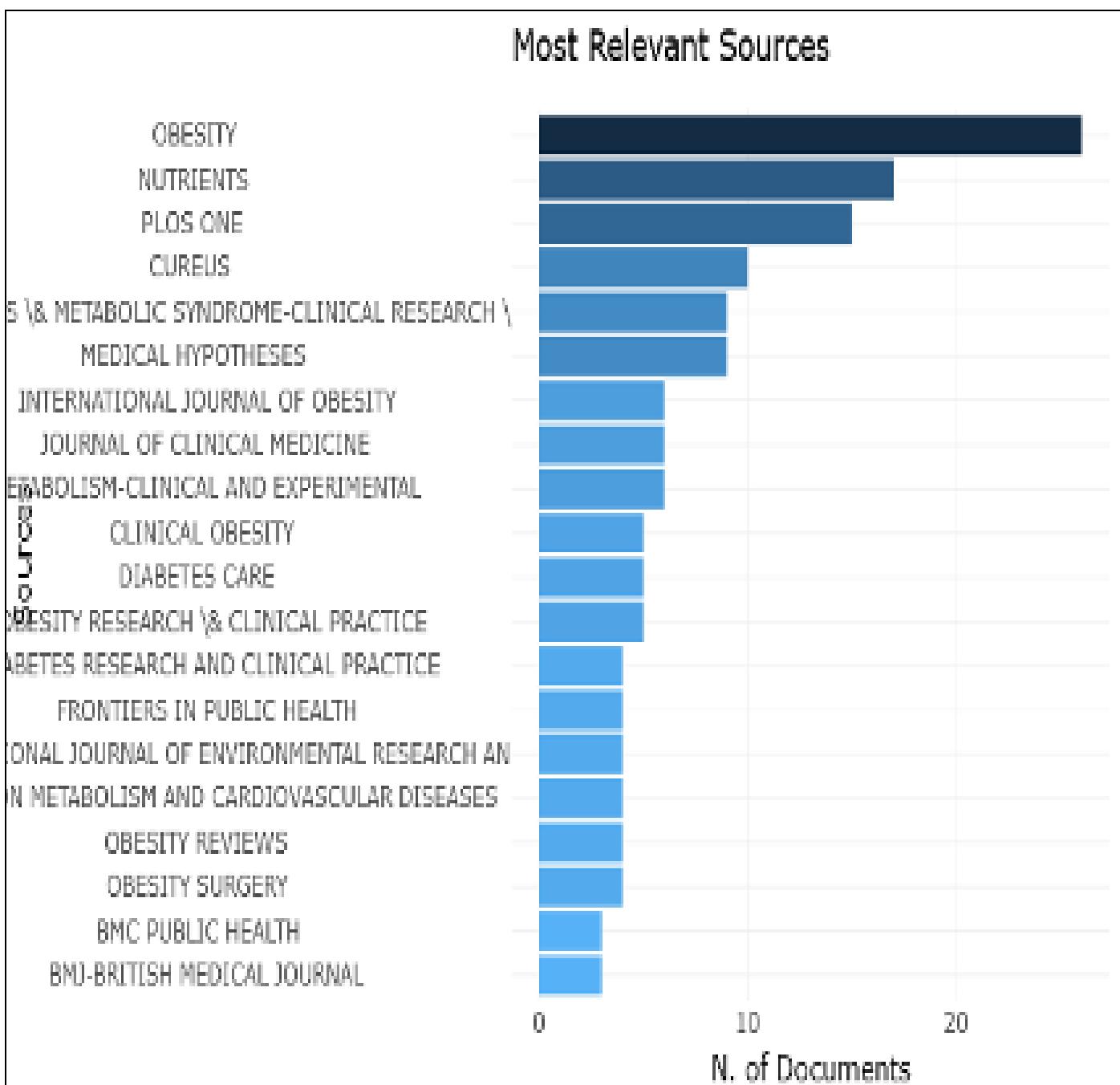


Figure 2. Most Relevant Sources

3.3 Most Cited Sources

Nowadays, research citation is an important aspect among academicians and scientists. So, Table 3 shows that Lancet is among the top-cited research journals (source), which is 3.62 %. Jam-J AM MED ASSOC has secured the second position among all published articles, 3.01 % related to obesity and coronavirus studies. New England journal of medicine has secured the third position and so on.

Table 3. Most Cited Journals Names (Sources)

Sr.#	Journals Names (Sources)	Articles	%
1	LANCET	687	3.621317
2	JAMA-J AM MED ASSOC	571	3.009857
3	NEW ENGL J MED	564	2.972959
4	OBESITY	414	2.182278
5	BMJ-BRIT MED J	206	1.085868
6	CLIN INFECT DIS	201	1.059512
7	PLOS ONE	197	1.038427
8	NUTRIENTS	171	0.901376
9	NATURE	148	0.780138
10	DIABETES CARE	140	0.737968
11	CIRCULATION	135	0.711612
12	METABOLISM	132	0.695799
13	LANCET RESP MED	124	0.653629
14	J MED VIROL	112	0.590375
15	MMWR-MORBID MORTAL W	96	0.506036
16	CELL	93	0.490222
17	OBES REV	92	0.484951
18	INTENS CARE MED	90	0.474408
19	LANCET INFECT DIS	89	0.469137
20	INT J OBESITY	88	0.463866

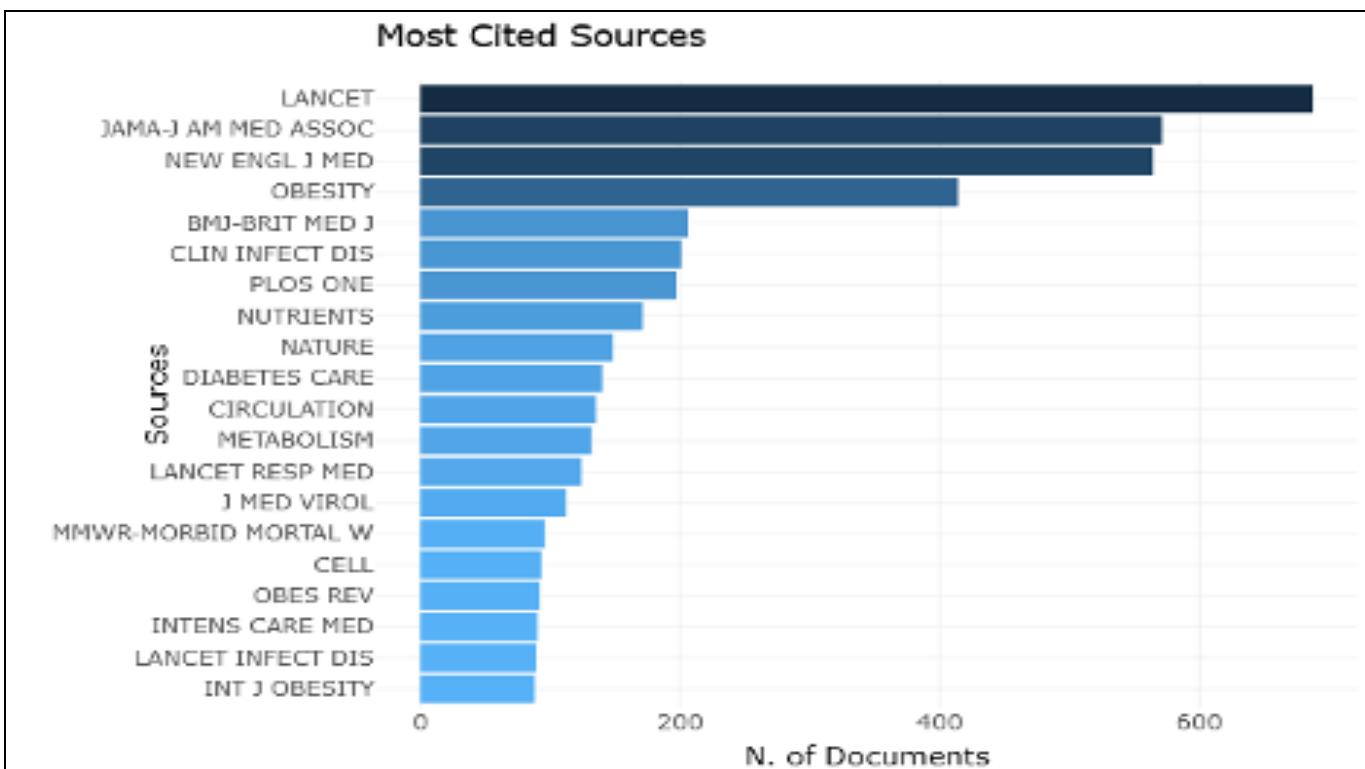


Figure 3. Most Cited Sources

3.4. Bradford Law

Bradford's law in Table 4 and figure 4, which is a pattern that calculates the exponentially decreasing returns of looking for references in science journals, was described by Samuel C. Bradford in 1934. Bradford's law divided all the sources (journals) into 3 zones (clusters), where cluster 1 consist 158 articles, which is 33.12 % among all the published research articles related to obesity and coronavirus (COVID19). On the other hand, cluster 2 consists of 162 (33.96 %) and cluster 3 have 157 (32.91 %) articles.

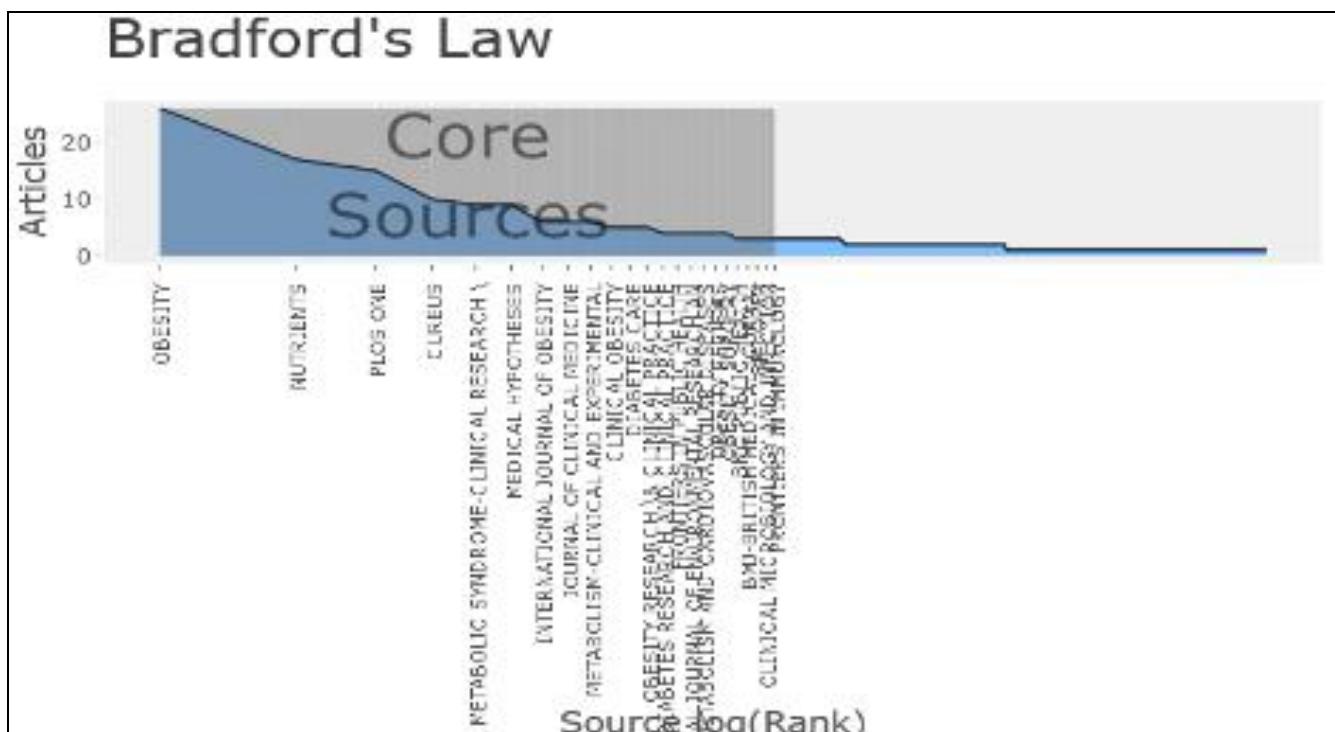


Figure 4. Bradford's Law

Table 4. Bradford's Law

Sr.#	Journals Names (Sources)	Rank	Freq	Cum. Freq	Zone
1	OBESITY	1	26	26	Zone 1
2	NUTRIENTS	2	17	43	Zone 1
3	PLOS ONE	3	15	58	Zone 1
4	CUREUS	4	10	68	Zone 1
5	DIABETES, METABOLIC SYNDROME CLINICAL RESEARCH & REVIEWS	5	9	77	Zone 1
6	MEDICAL HYPOTHESES	6	9	86	Zone 1
7	INTERNATIONAL JOURNAL OF OBESITY	7	6	92	Zone 1
8	JOURNAL OF CLINICAL MEDICINE	8	6	98	Zone 1
9	METABOLISM-CLINICAL AND EXPERIMENTAL	9	6	104	Zone 1
10	CLINICAL OBESITY	10	5	109	Zone 1
11	DIABETES CARE	11	5	114	Zone 1
12	OBESITY RESEARCH & CLINICAL PRACTICE	12	5	119	Zone 1
13	DIABETES RESEARCH AND CLINICAL PRACTICE	13	4	123	Zone 1
14	FRONTIERS IN PUBLIC HEALTH	14	4	127	Zone 1
15	INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	15	4	131	Zone 1
16	NUTRITION METABOLISM AND CARDIOVASCULAR DISEASES	16	4	135	Zone 1
17	OBESITY REVIEWS	17	4	139	Zone 1
18	OBESITY SURGERY	18	4	143	Zone 1
19	BMC PUBLIC HEALTH	19	3	146	Zone 1
20	BMJ-BRITISH MEDICAL JOURNAL	20	3	149	Zone 1

Top 50 Relevant Affiliations to produce COVID-19 and Obesity Scientific Literature have been analyzed. The most amazing facts have been found that among all the institutes of the world, Columbia University USA became first to produce covid19 and Obesity research which is 1.43 % overall.

Table 5. Most 50 Relevant Affiliations to produce COVID-19 and Obesity Scientific Literature

Sr.#	Affiliations	Articles	Per. (%)	Sr.#	Affiliations	Articles	Per. (%)
1	COLUMBIA UNIV	47	1.4321	26	UNIV PARIS	12	0.3656
2	SORBONNE UNIV	34	1.0360	27	KING SAUD UNIV	11	0.3352
3	HARVARD MED SCH	31	0.9445	28	UNIV OXFORD	11	0.3352
4	ICAHN SCH MED MT SINAI	28	0.8531	29	WEST VIRGINIA UNIV	11	0.3352
5	UNIV WASHINGTON	24	0.7313	30	CAIRO UNIV	10	0.3047
6	WENZHOU MED UNIV	23	0.7008	31	EMORY UNIV	10	0.3047
7	HUAZHONG UNIV SCI AND TECHNOL	22	0.6703	32	STANFORD UNIV	10	0.3047
8	ALBERT EINSTEIN COLL MED	20	0.6094	33	UNIV ROMA TOR VERGATA	10	0.3047
9	NATL AND KAPODISTRIAN UNIV ATHENS	20	0.6094	34	UNIV TORONTO	10	0.3047
10	UNIV SAO PAULO	20	0.6094	35	HARVARD TH CHAN SCH PUBL HLTH	9	0.2742
11	BOSTON UNIV	18	0.5484	36	IMPERIAL COLL LONDON	9	0.2742
12	UNIV LIVERPOOL	18	0.5484	37	JOHNS HOPKINS UNIV	9	0.2742
13	MASSACHUSETTS GEN HOSP	16	0.4875	38	UNIV CONNECTICUT	9	0.2742
14	UNIV CALIF IRVINE	15	0.4570	39	UNIV NAPLES FEDERICO II	9	0.2742
15	UNIV MILAN	15	0.4570	40	UNIV STRASBOURG	9	0.2742
16	UNIV PADUA	15	0.4570	41	ALICANTE GEN UNIV HOSP	8	0.2438
17	TULANE UNIV	14	0.4266	42	BEN GURION UNIV NEGEV	8	0.2438
18	UNIV COLORADO	14	0.4266	43	BROWN UNIV	8	0.2438
19	SAPIENZA UNIV ROME	13	0.3961	44	CEDARS SINAI MED CTR	8	0.2438
20	UNIV NACL AUTONOMA MEXICO	13	0.3961	45	FLORIDA STATE UNIV	8	0.2438
21	WASHINGTON UNIV	13	0.3961	46	HOSP CIVILS LYON	8	0.2438
22	INST NACL CIENCIAS MED AND NUTR SALVADOR ZUBIRAN	12	0.3656	47	INST SALUD CARLOS III	8	0.2438
23	TEL AVIV UNIV	12	0.3656	48	LONDON SCH HYG AND TROP MED	8	0.2438
24	UNIV BOLOGNA	12	0.3656	49	MED UNIV VIENNA	8	0.2438
25	UNIV MICHIGAN	12	0.3656	50	UNIV BRESCIA	8	0.2438

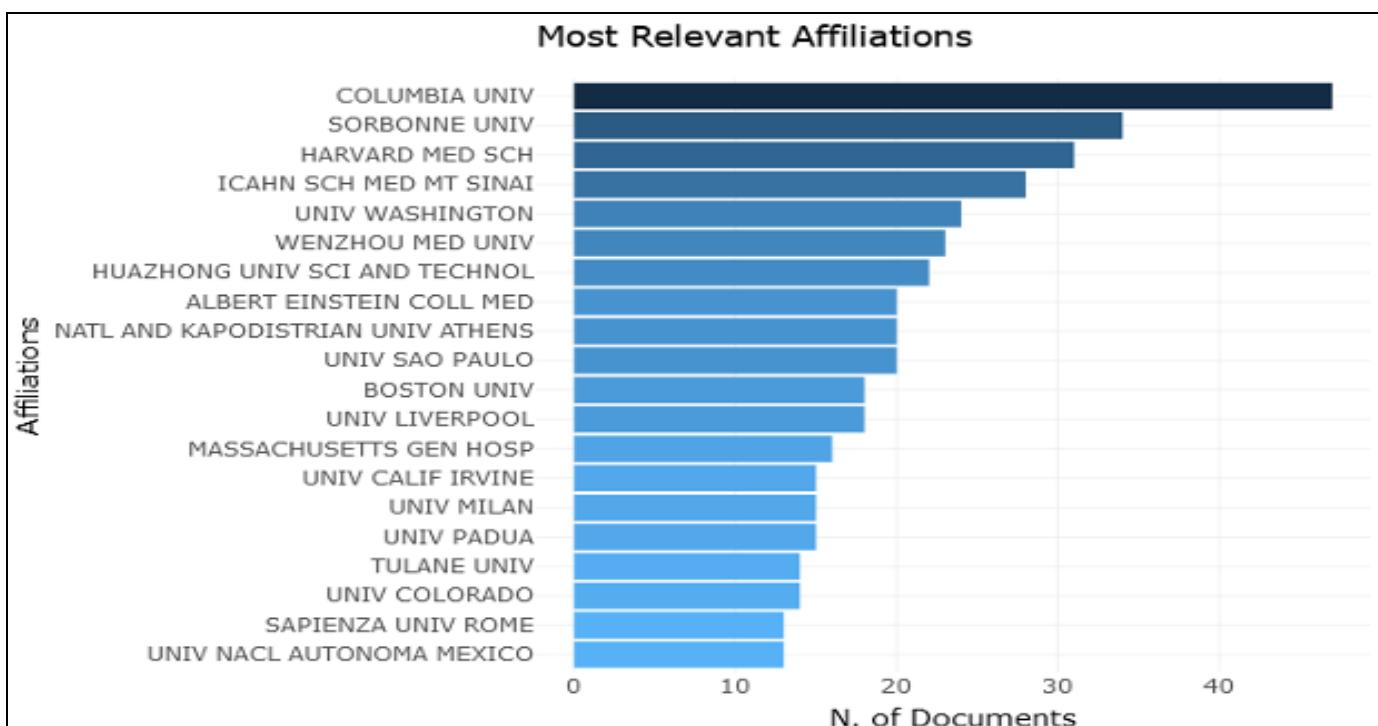


Figure 5 Most 20 Relevant Affiliations to produce COVID-19 and Obesity Scientific Literature

The top 20 Relevant Countries by corresponding Authors have been discussed in the following table. The world has the most influenced countries in table 6, where the USA has produced 32.63 % of relevant research articles related to COVID-19 and Obesity. Italy became second and the United Kingdom at third position and so on.

Table 6. Most Relevant Countries by corresponding Authors

Sr.#	Country	Articles	Freq.	SCP	MCP	MCP Ratio	% Article Freq
1	USA	155	0.32632	121	34	0.219	32.63158
2	ITALY	43	0.09053	36	7	0.163	9.052632
3	UNITED KINGDOM	40	0.08421	29	11	0.275	8.421053
4	FRANCE	30	0.06316	22	8	0.267	6.315789
5	SPAIN	25	0.05263	19	6	0.24	5.263158
6	BRAZIL	21	0.04421	18	3	0.143	4.421053
7	CHINA	20	0.04211	16	4	0.2	4.210526
8	MEXICO	20	0.04211	15	5	0.25	4.210526
9	GERMANY	9	0.01895	7	2	0.222	1.894737
10	NETHERLANDS	9	0.01895	6	3	0.333	1.894737
11	ISRAEL	8	0.01684	3	5	0.625	1.684211
12	CANADA	7	0.01474	3	4	0.571	1.473684
13	GREECE	7	0.01474	3	4	0.571	1.473684
14	POLAND	6	0.01263	5	1	0.167	1.263158
15	SAUDI ARABIA	5	0.01053	1	4	0.8	1.052632
16	TURKEY	5	0.01053	4	1	0.2	1.052632
17	AUSTRALIA	4	0.00842	1	3	0.75	0.842105
18	INDIA	4	0.00842	3	1	0.25	0.842105
19	KOREA	4	0.00842	4	0	0	0.842105
20	SWITZERLAND	4	0.00842	2	2	0.5	0.842105

Three main research groups are working together. The first cluster consists of 19 countries, 38 % contributors overall. It includes the USA, Germany, Turkey, Spain, Canada, India, other countries are displayed in table 7. The second cluster belongs to 6 countries, which is 12 % contributor overall, and it includes Denmark, Switzerland, Norway, Israel, Netherlands and Czech Republic. The third (11) cluster consists of 17 countries, with 34 % contributors overall. It includes Japan, Korea, Pakistan, Russia, China, and most of the European countries in the world, according to Table 7.

Table 7. Global Clusters of research collaborations on Obesity and Coronavirus

Clusters	1	2	11
% of Nodes	38%	12%	34%
Brazil		Denmark	Korea
France		Switzerland	Japan
United Kingdom		Netherlands	Egypt
USA		Norway	Pakistan
Italy		Israel	Russia
Turkey		Czech Republic	Hungary
Mexico			China
Germany			Australia
Ireland			Sweden
Spain			Austria
India			Poland
Saudi Arabia			Argentina
Romania			Philippines
Canada			Serbia
South Africa			Slovakia
Greece			Malaysia
Iran			Belgium
Lithuania			
Lebanon			

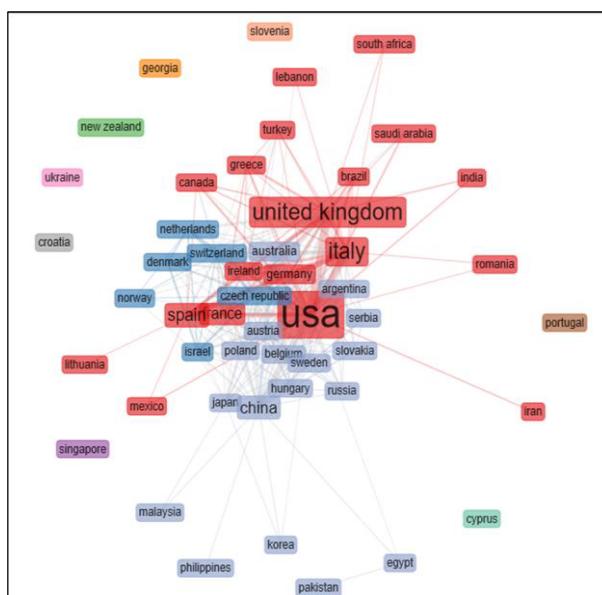


Figure 6. Most Relevant Countries by corresponding Authors

Table 8. Most Relevant Affiliations Map as research collaborations

Sr.#	Affiliations	Articles	%	Sr.#	Affiliations	Articles	%
1	COLUMBIA UNIV	47	1.43	26	UNIV PARIS	12	0.43
2	SORBONNE UNIV	34	1.05	27	KING SAUD UNIV	11	0.39
3	HARVARD MED SCH	31	0.97	28	UNIV OXFORD	11	0.40
4	ICAHN SCH MED MT SINAI	28	0.88	29	WEST VIRGINIA UNIV	11	0.40
5	UNIV WASHINGTON	24	0.76	30	CAIRO UNIV	10	0.36
6	WENZHOU MED UNIV	23	0.74	31	EMORY UNIV	10	0.36
7	HUAZHONG UNIV SCI AND TECHNOL	22	0.71	32	STANFORD UNIV	10	0.37
8	ALBERT EINSTEIN COLL MED	20	0.65	33	UNIV ROMA TOR VERGATA	10	0.37
9	NATL AND KAPODISTRIAN UNIV ATHENS	20	0.66	34	UNIV TORONTO	10	0.37
10	UNIV SAO PAULO	20	0.66	35	HARVARD TH CHAN SCH PUBL HLTH	9	0.33
11	BOSTON UNIV	18	0.60	36	IMPERIAL COLL LONDON	9	0.33
12	UNIV LIVERPOOL	18	0.60	37	JOHNS HOPKINS UNIV	9	0.33
13	MASSACHUSETTS GEN HOSP	16	0.54	38	UNIV CONNECTICUT	9	0.34
14	UNIV CALIF IRVINE	15	0.51	39	UNIV NAPLES FEDERICO II	9	0.34
15	UNIV MILAN	15	0.51	40	UNIV STRASBOURG	9	0.34
16	UNIV PADUA	15	0.51	41	ALICANTE GEN UNIV HOSP	8	0.30
17	TULANE UNIV	14	0.48	42	BEN GURION UNIV NEGEV	8	0.30
18	UNIV COLORADO	14	0.48	43	BROWN UNIV	8	0.30
19	SAPIENZA UNIV ROME	13	0.45	44	CEDARS SINAI MED CTR	8	0.30
20	UNIV NACL AUTONOMA MEXICO	13	0.45	45	FLORIDA STATE UNIV	8	0.31
21	WASHINGTON UNIV	13	0.45	46	HOSP CIVILS LYON	8	0.31
22	INST NACL CIENCIAS MED AND NUTR SALVADOR ZUBIRAN	12	0.42	47	INST SALUD CARLOS III	8	0.31
23	TEL AVIV UNIV	12	0.42	48	LONDON SCH HYG AND TROP MED	8	0.31
24	UNIV BOLOGNA	12	0.42	49	MED UNIV VIENNA	8	0.31
25	UNIV MICHIGAN	12	0.43	50	UNIV BRESCIA	8	0.31

Most Relevant Affiliations

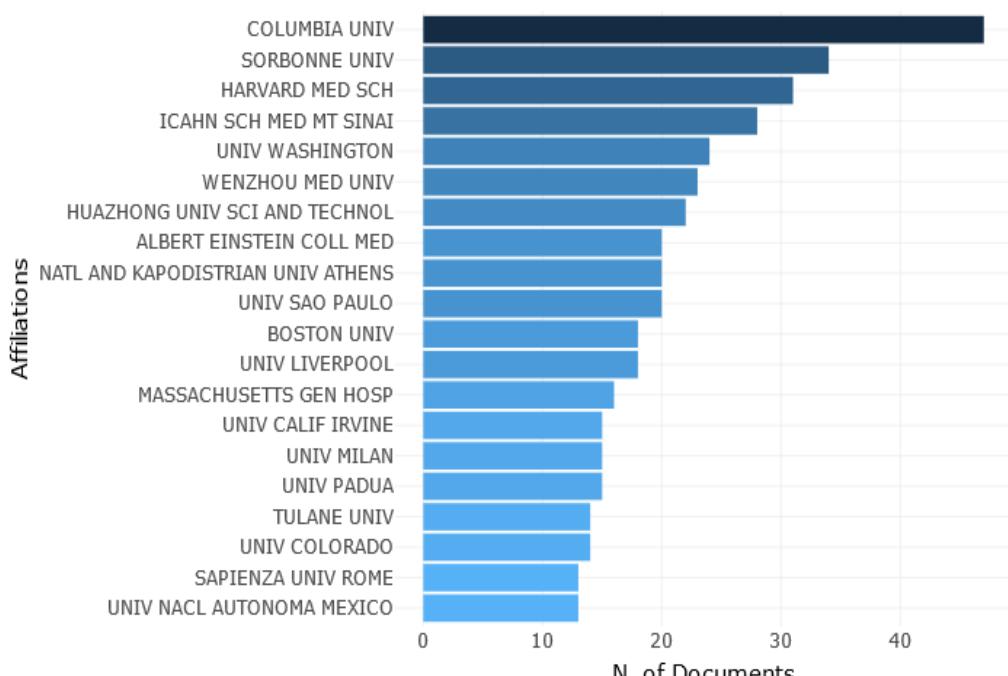


Table 9. Global Clusters of Most Relevant Affiliations on Obesity and Coronavirus

Clusters	1	2	3	21	22	23
% of Nodes	(8%)	(8%)	(16%)	(4%)	(16%)	(10%)
	Kings College London	Harvard medical school	Boston Univ.	Grp. Hosp St. Joseph	Sorbonne Univ.	Med. Univ. Vienna
	Institute Salud. Carlos iii	Stanford University	Univ. Modena and Reggio Emilia	Hosp. Civiles Lyon	Univ. Paris	Charles Prague univ.
	Hosp. Univ. Bellvitge	Icahn Sch. Med. Mt. Sinai	Univ. Cattolica Sacro Cuore		Hosp. Civils Lyon	Semmelweis Univ.
	Univ. Barcelona	Univ. Washington	Univ Milan		Ctr. Hosp. Agen. Nerac	Verona univ Hosp.
			Univ. Pisa		Ctr. Hosp. Poissy	Comenius Univ.
			Univ. Milano Bicocca		Inst. Mutualiste Montsouris	
			Univ. Padua		Inst. Arnault Tzanck	
			univ bologna		Natl and Kapodistrian Univ. Athens	

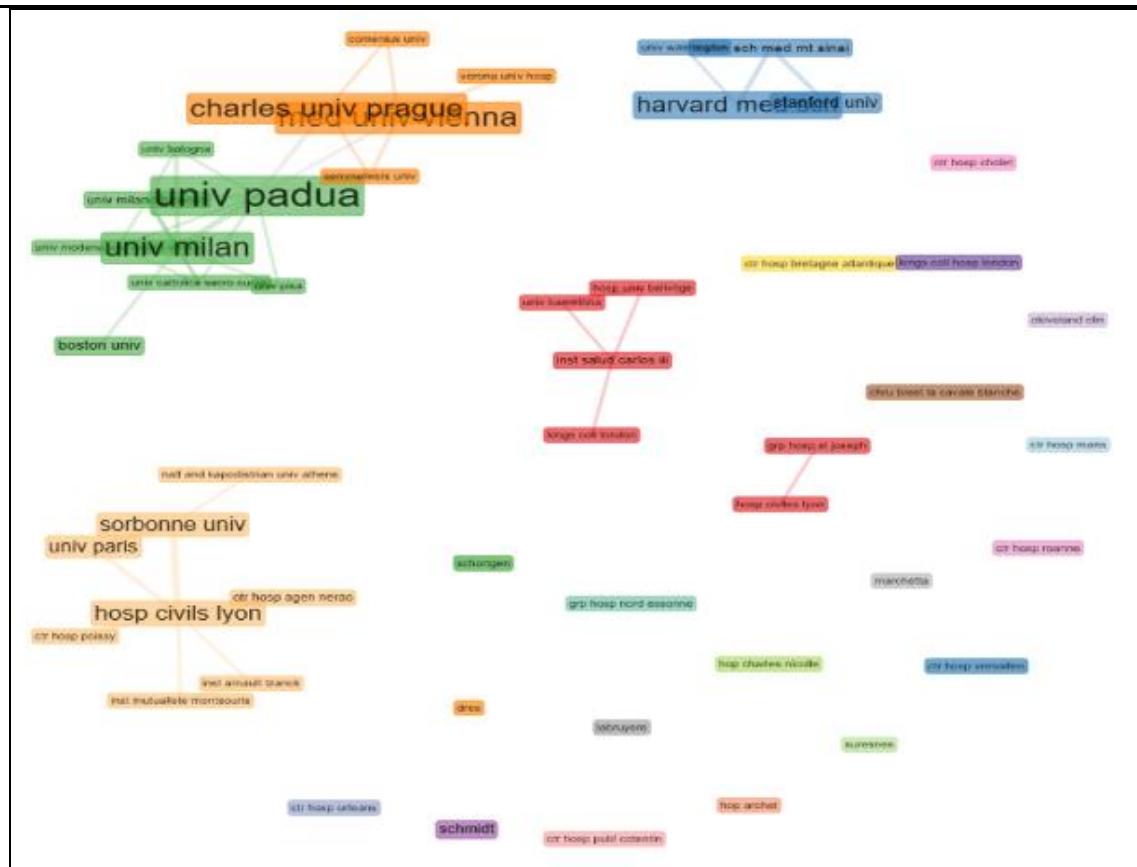


Table 10. Global GIS Mapping: Produced Scientific Literature on Obesity and COVID-19

Sr.#	Country	Count	%	Sr.#	Country	Count	%
1	USA	876	25.29	26	COLOMBIA	18	6.72
2	FRANCE	444	17.16	27	BELGIUM	16	6.40
3	ITALY	392	18.28	28	RUSSIA	14	5.98
4	UK	280	15.98	29	AUSTRIA	11	5.00
5	SPAIN	226	15.35	30	CZECH REPUBLIC	11	5.26
6	CHINA	150	12.04	31	SOUTH AFRICA	11	5.56
7	BRAZIL	118	10.77	32	SWEDEN	10	5.35
8	MEXICO	113	11.55	33	ARGENTINA	9	5.08
9	GERMANY	80	9.25	34	IRAN	9	5.36
10	ISRAEL	43	5.48	35	MOROCCO	9	5.66
11	GREECE	42	5.66	36	BULGARIA	8	5.33
12	NETHERLANDS	40	5.71	37	CHILE	8	5.63
13	VIETNAM	39	5.91	38	DENMARK	8	5.97
14	SWITZERLAND	38	6.12	39	HUNGARY	8	6.35
15	AUSTRALIA	37	6.35	40	ROMANIA	8	6.78
16	CANADA	37	6.78	41	BANGLADESH	7	6.36
17	INDIA	35	6.88	42	THAILAND	7	6.80
18	POLAND	31	6.54	43	INDONESIA	6	6.25
19	SAUDI ARABIA	31	7.00	44	IRAQ	5	5.56
20	SOUTH KOREA	30	7.28	45	LATVIA	5	5.88
21	JAPAN	24	6.28	46	PAKISTAN	5	6.25
22	SINGAPORE	24	6.70	47	PERU	5	6.67
23	EGYPT	23	6.89	48	KUWAIT	4	5.71
24	IRELAND	22	7.07	49	LITHUANIA	4	6.06
25	TURKEY	21	7.27	50	MALTA	4	6.45

Country Scientific Production

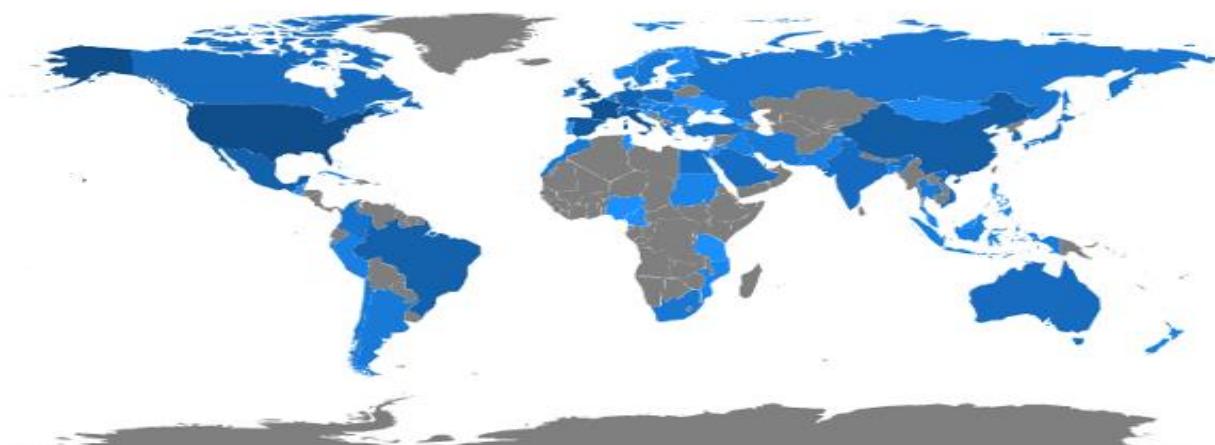


Figure 7. Global GIS Mapping: Produced Scientific Literature on Obesity and COVID-19

Table 11. Most Cited Countries WRT COVID-19 and Obesity

Sr.#	Country	Total Citations	Average Article Citations	% TC
1	USA	3366	21.72	56.21
2	UNITED KINGDOM	564	14.10	9.42
3	FRANCE	507	16.90	8.47
4	ITALY	462	10.74	7.72
5	SWITZERLAND	239	59.75	3.99
6	CHINA	189	9.45	3.16
7	SPAIN	127	5.08	2.12
8	GERMANY	63	7.00	1.05
9	MEXICO	58	2.90	0.97
10	GREECE	57	8.14	0.95
11	AUSTRALIA	46	11.50	0.77
12	BRAZIL	44	2.10	0.73
13	CZECH REPUBLIC	43	43.00	0.72
14	TURKEY	38	7.60	0.63
15	ISRAEL	34	4.25	0.57
16	NETHERLANDS	34	3.78	0.57
17	INDIA	28	7.00	0.47
18	POLAND	15	2.50	0.25
19	COLOMBIA	14	4.67	0.23
20	ROMANIA	14	7.00	0.23

Figure 8. Most Cited Countries WRT COVID-19 and Obesity

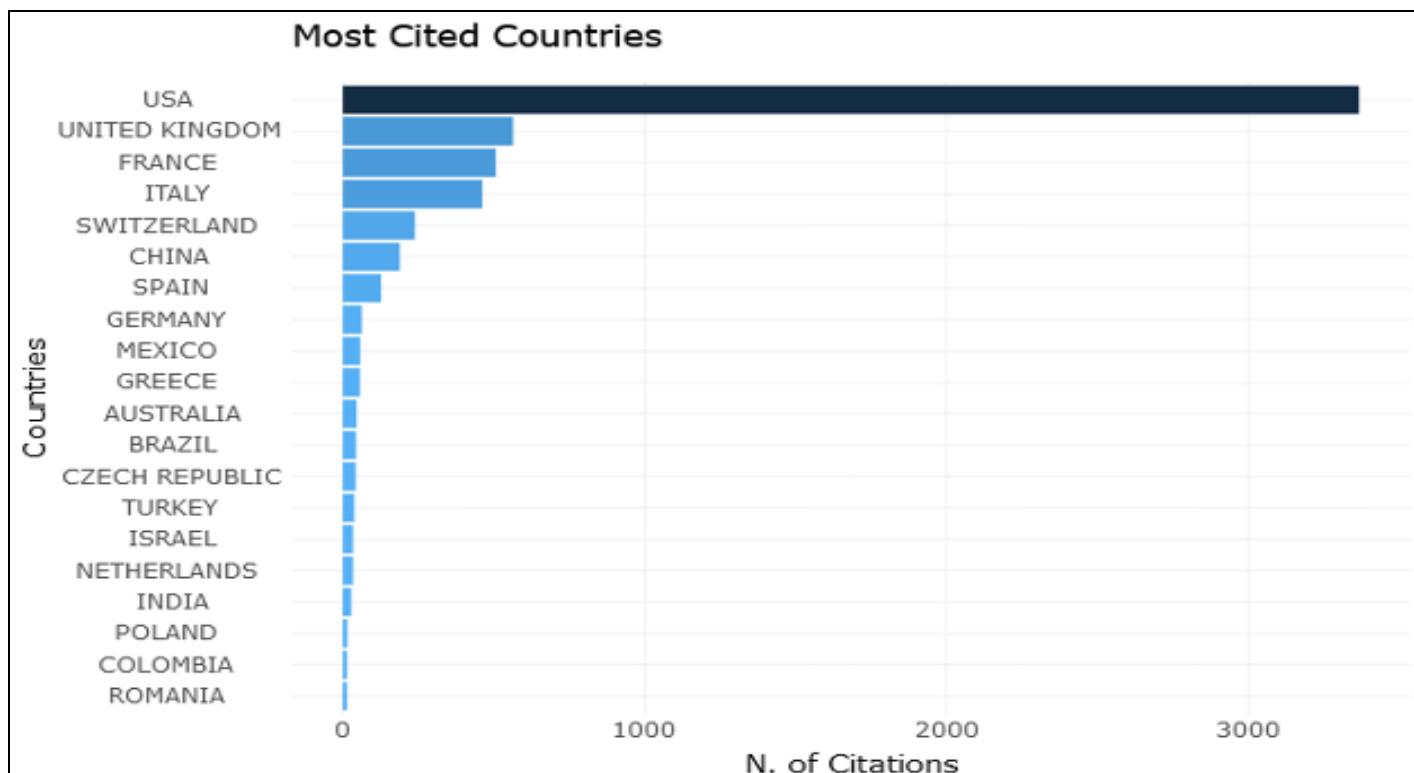


Table 12. Most Global Cited Documents

Sr.#	Paper	DOI	Total Citations	TC Year	per
1	RICHARDSON S, 2020, JAMA-J AM MED ASSOC	10.1001/jama.2020.6775	1394	697	
2	SIMONNET A, 2020, OBESITY	10.1002/oby.22831	402	201	
3	DOCHERTY AB, 2020, BMJ-BRITISH MEDICAL JOURNAL	10.1136/bmj.m1985 10.1016/S0140-6736(20)31189-2	290	145	
4	CUMMINGS MJ, 2020, LANCET		232	116	
5	MENTER T, 2020, HISTOPATHOLOGY	10.1111/his.14134 10.1016/S0140-6736(20)31187-9	195	97.5	
6	KUDERER NM, 2020, LANCET		181	90.5	
7	PRICE-HAYWOOD EG, 2020, N ENGL J MED	10.1056/NEJMsa2011686	149	74.5	
8	MUNIYAPPA R, 2020, AM J PHYSIOL -ENDOCRINOL METAB	10.1152/ajpendo.00124.2020	126	63	
9	KULCSAR KA, 2019, JCI INSIGHT	10.1172/jci.insight.131774	93	31	
10	PALAIODIMOS L, 2020, METAB -CLIN EXP	10.1016/j.metabol.2020.154262	92	46	
11	LUZI L, 2020, ACTA DIABETOL	10.1007/s00592-020-01522-8	81	40.5	
12	CAI Q, 2020, DIABETES CARE	10.2337/dc20-0576	79	39.5	
13	KALLIGEROS M, 2020, OBESITY	10.1002/oby.22859	78	39	
14	ARGENZIANO MG, 2020, BMJ-BRITISH MEDICAL JOURNAL	10.1136/bmj.m1996 10.1016/S2214-109X(20)30264-3	72	36	
15	CLARK A, 2020, LANCET GLOB HEALTH		65	32.5	
16	DI RENZO L, 2020, J TRANSL MED-a	10.1186/s12967-020-02399-5	64	32	
17	PUIG-DOMINGO M, 2020, ENDOCRINE	10.1007/s12020-020-02294-5	55	27.5	
18	CHEN Q, 2020, INFECTION	10.1007/s15010-020-01432-5	54	27	
19	PIETROBELLIA A, 2020, OBESITY	10.1002/oby.22861	52	26	
20	KRUGLIKOV IL, 2020, OBESITY	10.1002/oby.22856	51	25.5	

Figure 9. Most Global Cited Documents

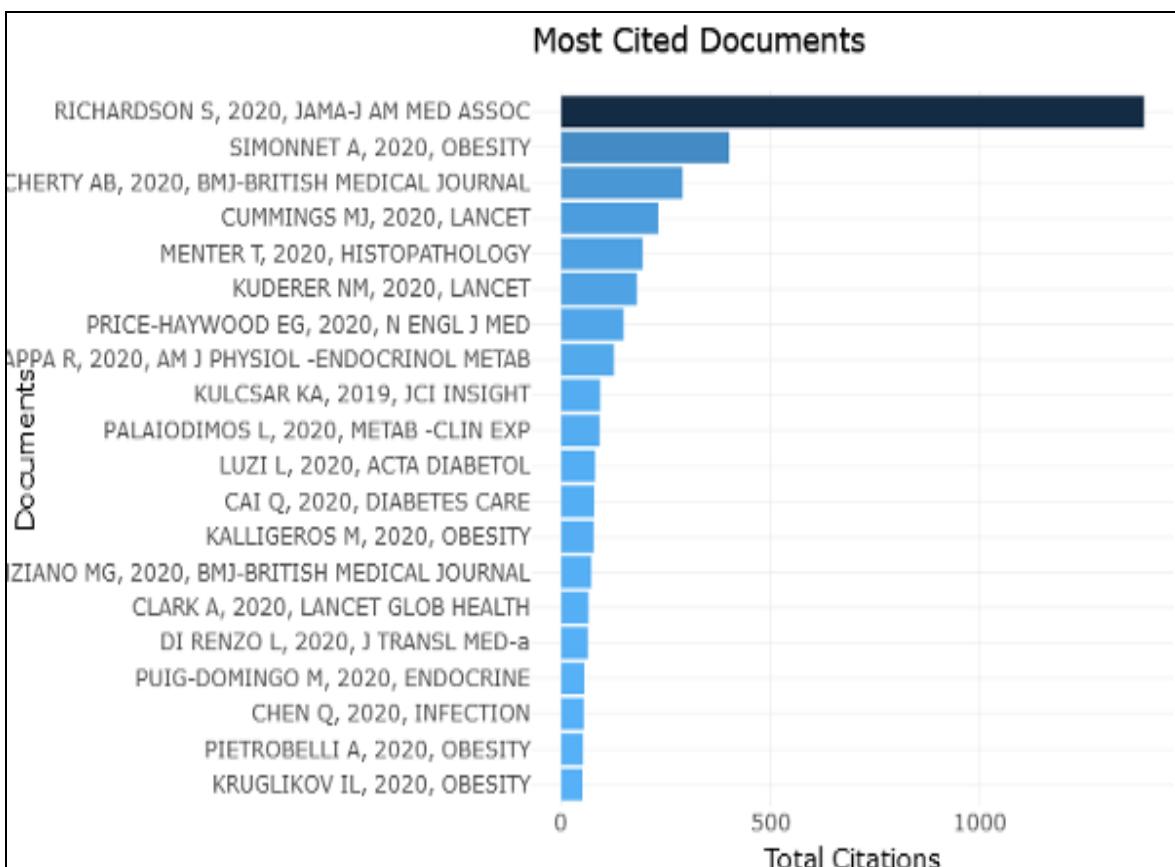


Table 13. Most Relevant Keywords

Sr.#	Keywords	Occurrences	%	Sr.#	Keywords	Occurrences	%
1	Covid-19	246	11.47	26	Covid	7	0.49
2	Obesity	105	5.53	27	Exercise	7	0.50
3	Sars-cov-2	76	4.24	28	Outcomes	7	0.50
4	Coronavirus	46	2.68	29	Stress	7	0.50
5	Pandemic	29	1.73	30	BMI	6	0.43
6	Diabetes	21	1.28	31	Comorbidity	6	0.43
7	Mortality	21	1.29	32	Coronavirus 2019 disease	6	0.44
8	Inflammation	15	0.94	33	Coronavirus 2019 (covid-19) disease	6	0.44
9	Physical activity	14	0.88	34	Diet	6	0.44
10	Ace2	13	0.83	35	Ethnicity	6	0.44
11	Risk factors	13	0.83	36	Health	6	0.44
12	Epidemiology	12	0.78	37	Sars-cov2	6	0.45
13	Diabetes mellitus	11	0.72	38	Syndrome	6	0.45
14	Pneumonia	11	0.72	39	Children	5	0.37
15	Hypertension	10	0.66	40	Complications	5	0.38
16	Quarantine	10	0.67	41	Coronavirus disease	5	0.38
17	Vitamin D	10	0.67	42	Cytokine storm	5	0.38
18	Bariatric surgery	9	0.61	43	Hospitalization	5	0.38
19	Body mass index	9	0.61	44	ICU	5	0.38
20	Comorbidities	8	0.55	45	Mental health	5	0.38
21	Lifestyle	8	0.55	46	Mexico	5	0.38
22	Lockdown	8	0.55	47	Nutrition	5	0.39
23	Overweight	8	0.56	48	Respiratory syndrome distress	5	0.39
24	Cardiovascular disease	7	0.49	49	Smoking	5	0.39
25	Coronavirus infections	7	0.49	50	Activity	4	0.31

Figure 10. Most Relevant Keywords

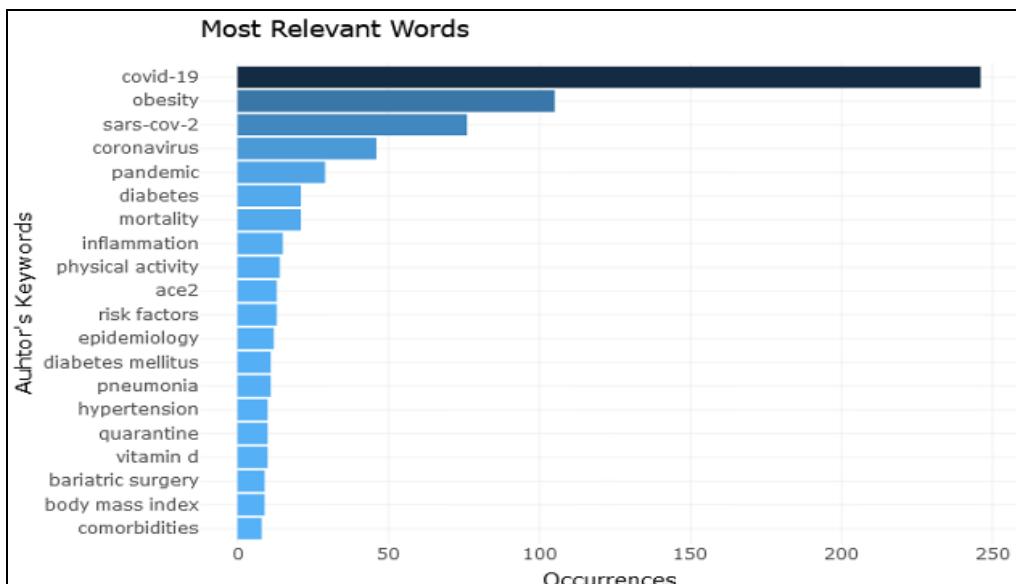


Figure 11. Most Relevant Keywords through Tree diagram

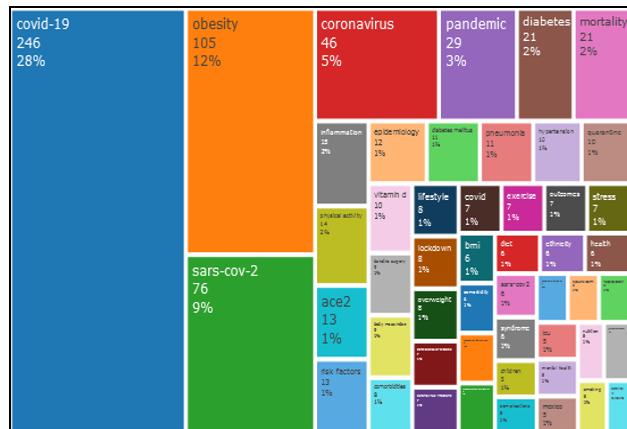


Figure 12. Comparison of Scientific Author Keywords and keywords Growth

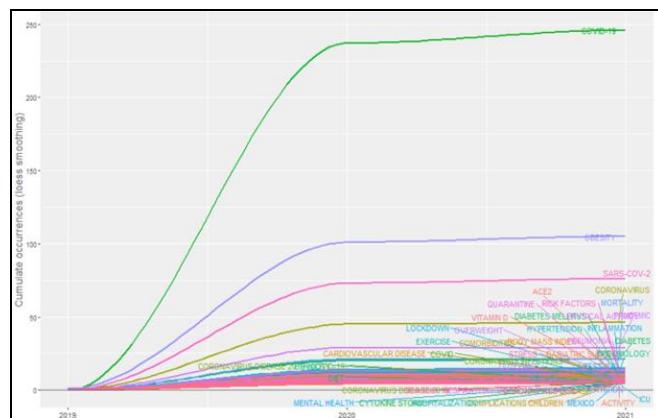


Figure 13. Comparison of Scientific Author Keywords through Dendrogram

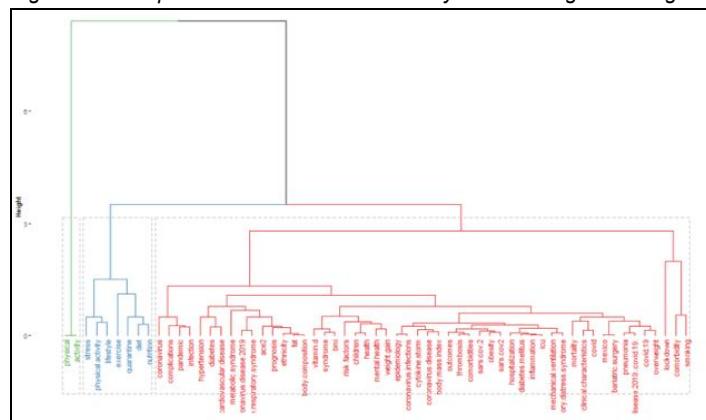


Figure 14. Institutes Co-occurrence Network with associations who have produced COVID-19 researches

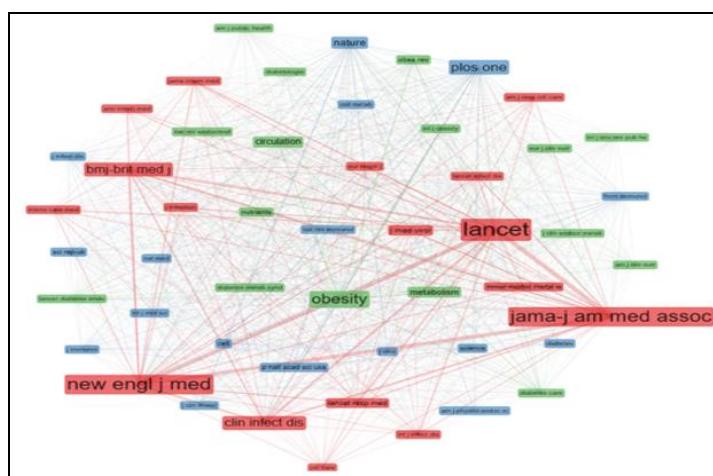


Figure 15. Country Collaboration Map

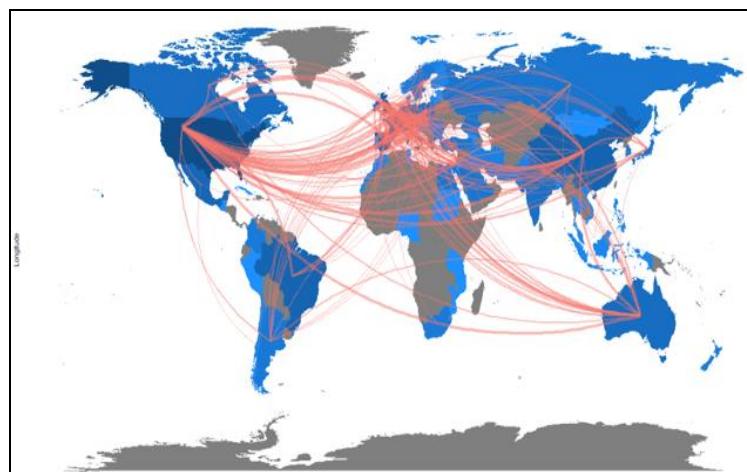
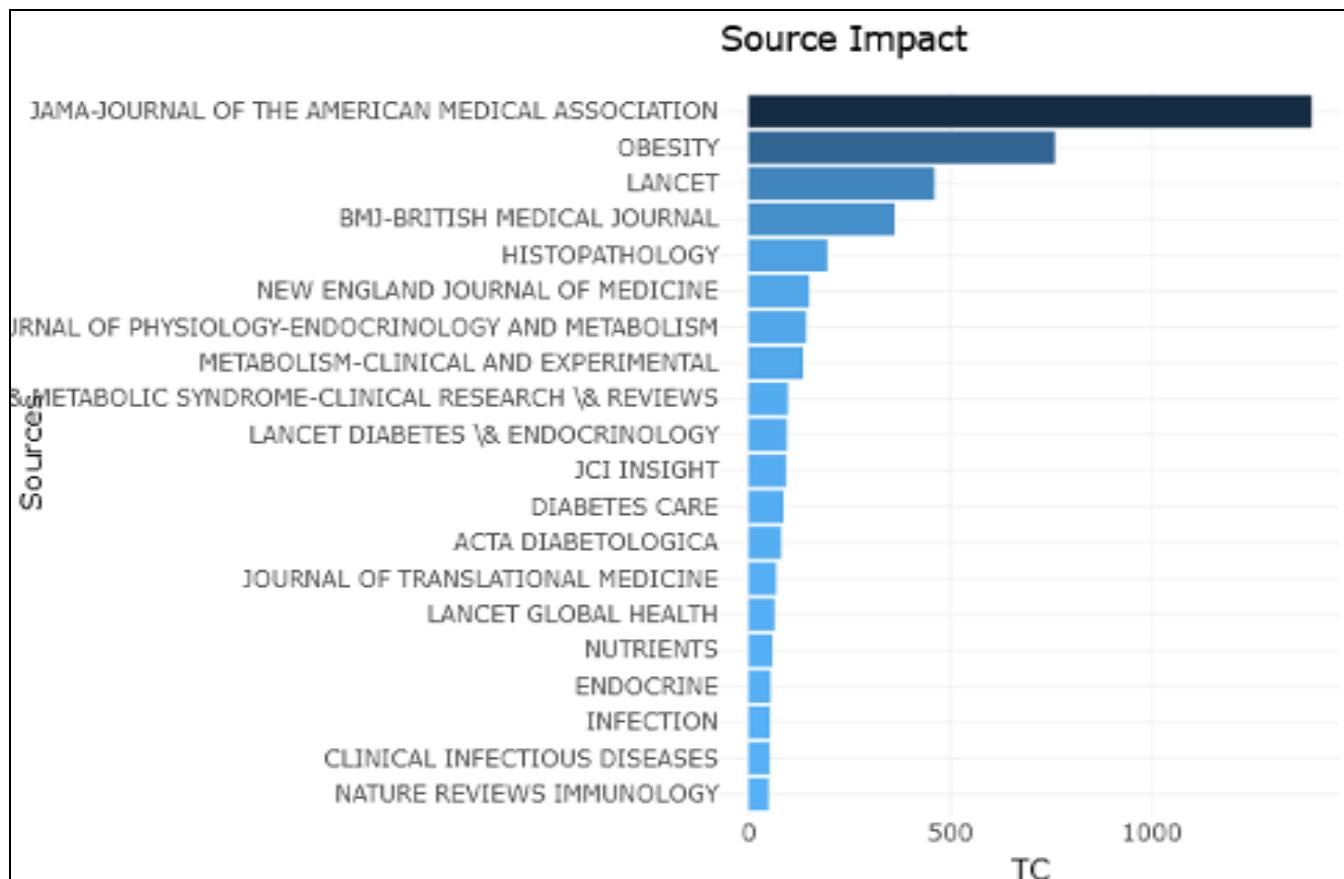


Table 14. Source Impact (using H Index and Total Citations)

Sr.#	Source	H index	G index	M index	Total Cited	NP	% Total Cited
1	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	1	1	0.5	1394	1	23.26
2	OBESITY	10	26	5	759	26	12.67
3	LANCET	3	3	1.5	460	3	7.68
4	BMJ-BRITISH MEDICAL JOURNAL	2	3	1	362	3	6.04
5	HISTOPATHOLOGY	1	1	0.5	195	1	3.25
6	NEW ENGLAND JOURNAL OF MEDICINE	1	2	0.5	150	2	2.50
7	AMERICAN JOURNAL OF PHYSIOLOGY-ENDOCRINOLOGY AND METABOLISM	2	2	1	143	2	2.39
8	METABOLISM-CLINICAL AND EXPERIMENTAL	5	6	2.5	135	6	2.25
9	DIABETES & METABOLIC SYNDROME-CLINICAL RESEARCH & REVIEWS	5	9	2.5	98	9	1.64
10	LANCET DIABETES & ENDOCRINOLOGY	3	3	1.5	95	3	1.59
11	JCI INSIGHT	1	1	0.33	93	1	1.55
12	DIABETES CARE	3	5	1.5	87	5	1.45
13	ACTA DIABETOLOGICA	1	1	0.5	81	1	1.35
14	JOURNAL OF TRANSLATIONAL MEDICINE	2	3	1	69	3	1.15
15	LANCET GLOBAL HEALTH	1	1	0.5	65	1	1.08
16	NUTRIENTS	4	7	2	59	17	0.98
17	ENDOCRINE	1	1	0.5	55	1	0.92
18	INFECTION	1	1	0.5	54	1	0.90
19	CLINICAL INFECTIOUS DISEASES	2	2	1	53	2	0.88
20	NATURE REVIEWS IMMUNOLOGY	2	2	1	50	2	0.83

Figure 16. Source Impact (using H Index and Total Citations)



4. Discussion

In this study, which we conducted on 477 articles that we obtained by filtering with PRISMA framework from approximately 93 thousand articles from the web of science database, we analyzed to what extent countries, publishers, universities and organizations focused on Covid-19 and Obesity and how effective they were in this field. In addition, we have observed inter-country collaborations and inter-institutional collaborations. The core journal's most relevant source is obesity with 29 articles followed by Nutrients Figure 2. Most Relevant Sources and Table 2. but the most cited sources are LANCET, JAMA (The Journal of the American Medical Association) and The New England Journal of Medicine, respectively (Fig. 3 and Table 3.) and obesity comes 4rd.

The top 3 universities that have contributed the most to the corpus are Columbia University, Sorbonne University and Harvard Medical School, respectively Fig. 8 and Table 5. According to the total citation and H index, JAMA: The Journal of the American Medical Association is the leading publisher with the highest source impact Fig. 18 and Table 16 OBESITY, LANCET, and BMJ also have very high source impact. The most cited publication in this field is the study of Richardson et al., showing the characteristics, comorbidities and outcomes of COVID-19 among 5700 hospitalized patients, which was published in JAMA in 2020 Table 9. The second most cited study is [10] on the prevalence of SARS-CoV-2 and Obesity published in OBESITY in 2020. Unsurprisingly, the most commonly used keywords in the articles we are working on are covid-19, obesity and coronavirus, respectively Table 10 and Figures 10-14. According to the authors' countries, in productivity, The USA ranks first with a big margin, followed by Italy and The United Kingdom Fig. 6 and Table 6.

Although studies emerged with cooperation between countries, most articles are single country publications Fig. 6. When we group the countries according to the Louvain clustering algorithm with 50 nodes and two minimum edges and opacity used as 0.7, America, China, Switzerland, and France fall in a cluster. The United Kingdom, Czech Republic, Germany, Israel and Ireland become another cluster. Finally, Italy, Spain and Australia are group 3. Countries contributing the most publications to COVID-19 and Obesity are the USA, France, Italy, UK and Spain, respectively Table 7 and Fig.7. The most cited countries by Covid-19 and obesity related publications are the United States, followed by the United Kingdom and France Table 8 and Fig. 8.

On the country collaboration map Fig. 15, we can see that The United States collaborated with China and Central Europe more than it did with other countries. These results show that developed countries are taking the lead on Covid-19 and Obesity research. This may be due to the advanced economy, having access to bigger data or advanced research environment and facilities. Cooperating more with developing and less developed countries can ensure more inclusive studies, that the studies are on a global scale and produce solutions that apply to the globe. Hence, we can overcome the pandemic and its effects worldwide.

As the governments advised, the only solution to the Covid-19 pandemic was to put people into quarantine or isolation. Humans are social animals, and restricting their social activities can cause anger, anxiety, frustrations and stress. These emotions can cause a change in lifestyle and eating behaviours, which is what happened during the Coronavirus pandemic. Change in the eating patterns and overall sleep-wake cycle (an unhealthy lifestyle) automatically led to obesity amongst the individuals [51].

This covid-19 pandemic has restricted individuals from staying indoors, making it difficult for overweight and obese people to exercise and stay healthy. The most vulnerable group are the obesity stricken, and during their lengthy stays inside the home, they are more prone to eating more, eating unhealthy, being inactive and lazy. All these factors together have a strong association with obesity. Gaining weight due to this sedentary lifestyle has led to low-self esteem in obese individuals causing depression and anxiety in return leading to eating disorders resulting again in Obesity [52].

Obesity is caused by a low level of physical activity, putting individuals at health risk. People confined at homes were forced to limit their outdoor activities, increasing the risk of obesity and related metabolic diseases. Altogether this led to an increase in healthcare costs. This situation is not confined to some specific area but is seen all around the globe. Amongst all the covid patients admitted, 42% of the population was affected with obesity making it the second highlighted comorbidity [53].

Covid-19 has caused disease and has led to unemployment, loneliness, and domestic violence. These situations affect individuals psychologically, linked with being overweight and obese [54]. Another research from the Journal of Obesity stated that in a study on young people's lifestyle behaviour, consumption of fruit intake increased (more fructose, more chances of gaining weight). Other food items that were increasingly consumed were potato chips, red meat and sugary things [55]. All these food items are rich in fat and sugars, contributing to weight gain and obesity. Excessive amounts of refined food products (with high carbohydrates) lead to increased insulin resistance and fat and extra nutrients storage into fatty tissues. High fat and sugary foods also are addictive, and individuals tend to overeat them in return leading to Obesity [56].

How does one get obese? We get obese by consuming energy-dense food items, higher caloric intake, and decreased physical activity levels. People followed all these obesity causing factors because of the quarantine imposed by the government. Individuals could not go to the gym to work out, children were unable to go to parks to play, and workers were sitting in their homes in front of desks to work online (56). This period of inactivity and consumption of increased food caused obesity rates to rise.

Obesity is considered a major health risk, and it puts you at risk of COVID-19. AF AbdelMassih, in his article on the genetic basis of COVID-19, tells us that the reason obesity can lead to Covid-19 is the suppression of T cells, depletion of memory B cells and dominance of macrophages M1. Along with all this, there is an increase in overall inflammation putting the human body at risk. He identifies single-cell sequencing as an appropriate means to identify individuals at risk of COVID-19 due to Obesity [57].

During the Coronavirus pandemic, a survey study of the supermarkets showed that shelves containing staple foods, i.e. flour, rice, beans, and others, were full. Contrary to this, shelves containing snack items like crisps, sugary beverages and ready-made food items were emptier [58]. This shows a rise in the eating of processed and ready to eat foods which cause excessive weight gain and obesity. If this unhealthy eating and lifestyle are not transformed into a healthy one, the Covid-19 pandemic might progress towards an obesity pandemic [59].

5. Conclusion and Future Research Direction

Coronavirus, which rapidly surrounded the world and caused the pandemic, triggered many side diseases along with the COVID-19 disease. This study tried to provide informative guidance for decision-makers to conclude the efficient way of rethinking the global clustering of big think tanks. The study presented our findings with a series of figures and tables with the expectation to be useful for researchers. The findings contain important information about the articles, countries, and publishers, including their worldwide impact and interrelationships.

Our study aimed to highlight the growing issue of "Obesity" due to this COVID-19 pandemic with people confined at homes; limited physical activity combined with an unhealthy lifestyle. Through bibliometric analysis, the study presented that the people have increased the number of weight scales and a rise in body fat levels in most individuals during the Coronavirus period. There is a strong association with global human obesity by getting bibliometric results. Dietary equilibrium and execution of valuable food sources should be expected to confront this conceivable COVID-19 pandemic.

This study was conducted on a limited amount of data since only the WoS database is used, but future studies should be extended by including publications from different databases. The study can be further detailed by adding a systematic literature review. The following article may be used for policymakers and practitioners. In addition, different metrics can be included. For instance, historical change can be analyzed. A separate and more detailed analysis can be carried out on keywords.

6. References

- [1]. WHO. World Health Organization. Pneumonia of unknown cause – China. Disease outbreak news: 5 January 2020.
- [2]. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8).
- [3]. Gorbatenko AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Vol. 5, *Nature Microbiology.* 2020.
- [4]. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med.* 2020;382(13).
- [5]. WHO. WHO Timeline - COVID-19. Wold Health Organization. 2020.
- [6]. World Health Organization. World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020.
- [7]. Zhu J, Zhong Z, Ji P, Li H, Li B, Pang J, et al. Clinicopathological characteristics of 8697 patients with COVID-19 in China: A meta-analysis. *Fam Med Community Heal.* 2020;8(2).
- [8]. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395(10229).
- [9]. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes.* 2008;32(9).
- [10]. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, et al. High Prevalence of Obesity in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) Requiring Invasive Mechanical Ventilation. *Obesity.* 2020;28(7).
- [11]. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected with SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA - J Am Med Assoc.* 2020;323(16).
- [12]. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med.* 2020;382(18).
- [13]. Hou YJ, Okuda K, Edwards CE, Martinez DR, Asakura T, Dinnon KH, et al. SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell.* 2020;182(2).
- [14]. Le Maréchal M, Morand P, Epaulard O, Némoz B. COVID-19 in clinical practice: A narrative synthesis. Vol. 50, *Medecine et Maladies Infectieuses.* 2020.
- [15]. Mohammad S, Aziz R, Al Mahri S, Malik SS, Haji E, Khan AH, et al. Obesity and COVID-19: what makes obese host so vulnerable? Vol. 18, *Immunity and Ageing.* 2021.
- [16]. WHO. WHO. World Health Organization (WHO): Obesity and overweight. World Health Organization. 2020.
- [17]. Rössner S. Obesity: The disease of the twenty-first century. *Int J Obes.* 2002;
- [18]. Garfield E. Historiographic mapping of knowledge domains literature. *Journal of Information Science.* 2004.
- [19]. Bettencourt LMA, Kaur J. Evolution and structure of sustainability science. *Proc Natl Acad Sci U S A.* 2011;
- [20]. Golan MS, Jernigan LH, Linkov I. Trends and applications of resilience analytics in supply chain modeling: systematic literature review in the context of the COVID-19 pandemic. *Environ Syst Decis [Internet].* 2020;40(2):222–43. Available from: <https://doi.org/10.1007/s10669-020-09777-w>
- [21]. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: Systematic review of early scientific literature. *J Med Internet Res.* 2020;22(11).
- [22]. Odore A, Salvati S, Bellini L, Bucci D, Capraro M, Gaetti G, et al. The runaway science: a bibliometric analysis of the COVID-19 scientific literature. *Acta Biomed.* 2020;91(5):34–9.

- [23]. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Phys Ther.* 2009;89(9):873–80.
- [24]. Boyack KW, Klavans R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *J Am Soc Inf Sci Technol.* 2010;
- [25]. Zitt M, Bassecoulard E. Development of a method for detection and trend analysis of research fronts built by lexical or cocitation analysis. *Scientometrics.* 1994;
- [26]. Gmür M. Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics.* 2003.
- [27]. De Bellis N. Bibliometrics and sisCitation Analy. *Journal of the American Society for Information Science and Technology.* 2009.
- [28]. Fetscherin M, Heinrich D. Consumer brand relationships research: A bibliometric citation meta-analysis. *J Bus Res.* 2015;68(2).
- [29]. Albanez JAP, Garcia SFA, Galli LCA. Métodos de Pesquisa de Satisfação de Clientes: um Estudo Bibliométrico. *Rev Bras Pesqui Mark Opinião e Mídia.* 2015;16.
- [30]. Verma S, Gustafsson A. Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. *J Bus Res.* 2020;118.
- [31]. Maalouf FT, Mdawar B, Meho LI, Akl EA. Mental health research in response to the COVID-19, Ebola, and H1N1 outbreaks: A comparative bibliometric analysis. *J Psychiatr Res.* 2021;132.
- [32]. Garfield E. Citation indexing for studying science. *Nature.* 1970;227(5259).
- [33]. Tamara A, Tahapary DL. Obesity as a predictor for a poor prognosis of COVID-19: A systematic review. *Diabetes Metab Syndr Clin Res Rev.* 2020;14(4).
- [34]. Groos O V., Pritchard A. Documentation notes. Vol. 25, *Journal of Documentation.* 1969.
- [35]. Nowakowska J, Sobocińska J, Lewicki M, Lemańska Ż, Rzymski P. When science goes viral: The research response during three months of the COVID-19 outbreak. *Biomed Pharmacother.* 2020;129.
- [36]. Van Leeuwen T. The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. In: *Scientometrics.* 2006.
- [37]. Dandona P, Aljada A, Chaudhuri A, Mohanty P, Garg R. Metabolic syndrome: A comprehensive perspective based on interactions between obesity, diabetes, and inflammation. *Circulation.* 2005;111(11):1448–54.
- [38]. Van Raan AFJ. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics.* 2005.
- [39]. Small H. Co-citation in the scientific literature: A new measure of the relationship between two documents. *J Am Soc Inf Sci.* 1973;
- [40]. Koondhar MA, Shahbaz M, Memon KA, Ozturk I, Kong R. A visualization review analysis of the last two decades for environmental Kuznets curve “EKC” based on co-citation analysis theory and pathfinder network scaling algorithms. *Environ Sci Pollut Res.* 2021;28(13).
- [41]. Zyoud SH, Zyoud AH. Coronavirus disease-19 in environmental fields: a bibliometric and visualization mapping analysis. *Environ Dev Sustain.* 2021;23(6).
- [42]. Chahrou M, Assi S, Bejjani M, Nasrallah AA, Salhab H, Fares MY, et al. A Bibliometric Analysis of COVID-19 Research Activity: A Call for Increased Output. *Cureus.* 2020;
- [43]. Murillo J, Villegas LM, Ulloa-Murillo LM, Rodríguez AR. Recent trends on omics and bioinformatics approaches to study SARS-CoV-2: A bibliometric analysis and mini-review. *Comput Biol Med.* 2021;128.
- [44]. Földi M, Farkas N, Kiss S, Zádori N, Váncsa S, Szakó L, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. Vol. 21, *Obesity Reviews.* 2020.
- [45]. Khan N, Qureshi MI. The outbreak of novel coronavirus (2019-COVID) Pandemic, its impacts on vaccination development and health care. A systematic literature review paper. *Eur J Mol Clin Med.* 2020;7(3):2845–70.
- [46]. Chen H, Yang Y, Yang Y, Jiang W, Zhou J. A bibliometric investigation of life cycle assessment research in the web of science databases. *Int J Life Cycle Assess.* 2014;19(10):1674–85.
- [47]. Broadus RN. Toward a definition of ‘bibliometrics’. *Scientometrics.* 1987;
- [48]. Zhai F, Zhai Y, Cong C, Song T, Xiang R, Feng T, et al. Research progress of coronavirus based on bibliometric analysis. *Int J Environ Res Public Health.* 2020;17(11).
- [49]. Abarca VMG, Palos-Sánchez PR, Rus-Arias E. Working in Virtual Teams: A Systematic Literature Review and a Bibliometric Analysis. *IEEE Access.* 2020;8:168923–40.
- [50]. Ho JSY, Fernando DI, Chan MY, Sia CH. Obesity in COVID-19: A Systematic Review and Meta-analysis. *Ann Acad Med Singapore.* 2020;49(12).
- [51]. Mattioli AV, Pinti M, Farinetti A, Nasi M. Obesity risk during collective quarantine for the COVID-19 epidemic. Vol. 20, *Obesity Medicine.* 2020.
- [52]. Bhasker AG, Greve JW. Are Patients Suffering from Severe Obesity Getting a Raw Deal During COVID-19 Pandemic? Vol. 30, *Obesity Surgery.* 2020.
- [53]. Siddiqui SA, Jakaria M. Lockdown leading obesity and its possible impacts on the second wave of COVID-19. Vol. 19, *Bangladesh Journal of Medical Science.* 2020.

- [54]. Hamer M, Stamatakis E. Inflammation as an intermediate pathway in the association between psychosocial stress and obesity. *Physiol Behav*. 2008;94(4).
- [55]. Pietrobelli A, Pecoraro L, Ferruzzi A, Heo M, Faith M, Zoller T, et al. Effects of COVID-19 Lockdown on Lifestyle Behaviors in Children with Obesity Living in Verona, Italy: A Longitudinal Study. *Obesity*. 2020;28(8).
- [56]. Abbas AM, Fathy SK, Fawzy AT, Salem AS, Shawky MS. The mutual effects of COVID-19 and obesity. Vol. 19, *Obesity Medicine*. 2020.
- [57]. AbdelMassih AF, Fouada R, Kamel A, Mishriky F, Ismail HA, El Qadi L, et al. Single cell sequencing unravelling genetic basis of severe COVID19 in obesity. Vol. 20, *Obesity Medicine*. 2020.
- [58]. Rundle AG, Park Y, Herbstman JB, Kinsey EW, Wang YC. COVID-19–Related School Closings and Risk of Weight Gain Among Children. Vol. 28, *Obesity*. 2020.
- [59]. Stefan N, Birkenfeld AL, Schulze MB. Global pandemics interconnected — obesity, impaired metabolic health and COVID-19. Vol. 17, *Nature Reviews Endocrinology*. 2021.



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:
Asad_khawaja et al. (2024). "Auditory Illusions: Detecting Synthetic Speech". FUSST International Conference on Computer Science and Engineering



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024 Asad_khawaja et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

Auditory Illusions: Detecting Synthetic Speech

Asad Khawaja ^a Suleman Ahmed ^a Muhammad Taha Farooq ^a Aania Majeed ^b

^a Department of Software Engineering, Foundation University, Islamabad, Rawalpindi, Pakistan.

^b Department of Computer Software Engineering, Military College of Signals, NUST, Pakistan.

* Corresponding author: asadkhawaja2345@gmail.com

Abstract:

Deepfake voice, synthetic voice, or voice cloning, employs artificial intelligence to forge a clone of someone's voice. This technology has reached the point where it can reproduce a human voice accurately in likeness and tone. This study concentrates on designing a machine-learning model to detect deepfake audio, handling the increasing abuse of voice cloning technology in several areas. ASVspoof 2019 dataset was used to transform audio samplings into spectrograms, which were then fed to a Convolutional Neural Network classifier. The model was developed to differentiate real, voice conversion, and text-to-speech audio files. Critical findings documented a high accuracy rate, with the model attaining 97-98% training accuracy and 99% accuracy on the test data. The study emphasizes the prospect of CNNs to enhance audio security by accurately identifying synthetic audio, for mitigation of the risks associated with voice cloning technology.

Keywords: Deepfake detection; Machine Learning; Neural Network.

1. Introduction

Voice cloning technology has developed over the years and is being used in various fields, including security and biometric identification, health, entertainment, and games. As the application of this kind of technology rises there emerges the concern of security risks that accompany such weakness [1].

Besides voice cloning, computer-generated audio has evolved and is now indistinguishable from any natural or real speech. This leads the world to a new host of scams that are committed by fraudsters who seek to use innocent people to achieve their goals. A swift response to this problem is needed in the security field [2].

Prior works of this kind rely on features such as the (MFCC), (CQCC) and (ENF) among others to generate numerical representations of audio data and then utilize the resultant data to train a model which is the (SVM) and the (HMM). [3][4][5]. This is a problem we're trying to solve by creating a model that can recognize whether the audio origin is real or else synthesized with the help of a machine learning algorithm. To this effect, we employ (CNN) classifier model. CNN is a Deep neural learning technique in which an endeavor is made to mimic the learning mechanism of the human brain to train a model. It operates to process images as the input. In this technique, Input data takes the form of images on which model parameters are placed and tested. As a first step, the audio samples of a given dataset are converted into spectrograms before feeding them into the CNN model in a training fashion. We then feed the above-detected files into CNN to segregate them as real and fake audio files. We tuned the parameters of the model to enhance its performance. For the assessment of the performance of the model, we provide accuracy, recall, and precision of the model.

1.1. Research Objectives

The research objectives are as follows and are significant to the development of a reliable machine learning system for distinguishing between original audio and a fake one. The objectives contribute to the improvement of the model performance and guarantee the effectiveness of generalization about audio verification and forensic analysis.

1.1.1. Research Objective 1

Thus, the most important specific goal is to develop and train a machine learning model for the classification of faked or deepfake audio content with high accuracy and low loss. This entails tuning the model's parameters to be able to predict well, reduce overfitting, and increase the performance of the audio verification system.

1.1.2. Research Objective 2

The second goal is to evaluate instructions and validations of the model and improve the model's generalization capacity. This will be beneficial in detecting overfitting and thereby, enhance the ability of the model to classify other features in the audio, which will assist in improving the field of audio forensics and enhancing the efficiency of the detection systems.

1.2. Research Contributions

The value of this research is primarily in the improvement of the verification methods for the audio signal and, most importantly, to counter the increasing cases of audio manipulation and deepfakes. In principle, it approximates the ideas of audio genuineness, defines some concepts to be useful in the case of the presence of fake audio, and generalizes theoretical dependencies that may describe links between the characteristics of audio and its legitimacy. In the empirical context, the study assesses these theoretical relations, decides upon the moderating variables that include noise, analyzes how some dimensions of audio moderate authenticity categorization, and evaluates the psychometric qualities of verification measures. The effectiveness of the proposed method has been shown and it has been proved that it has a higher performance than the existing methods of detecting deepfake audio. These contributions build upon and expand the current human and mechanistic understanding of manipulated audio and contribute to both the academic and objective analysis of audio forensics.

2. Related Work

Deep fake voice technology is on the rise and this has given rise to more issues regarding deceit. The focus of this review is to discuss the developments in methods to detect synthetic voice manipulations in the last few years and to assess how effective they are.

The current research paper by Dhamyal, H. , Ali, A. , Qazi, I. , & Raza, A , (2021) discusses how two micro features VOT and coarticulation can be used to differentiate between the bonafide audio and the synthesized one, and how these micro features can be used for building light-weight models for fake audio detection, particularly in low resource environments. The existence of possible variations concerning the phonemes assumed in the samples of audio may impact VOT definition as well as formant analysis. That is, the AutoVOT system which was used to detect VOT has an average error that varies from 5 ms which is smaller than the range of VOTs observed [7].

In extending the audio analysis techniques, (Tami T. , & Lars T. , 2021) reveal that although Fourier transforms have the capacity to differentiate deep fake voices from real voices, the 2D CNN model has shown greater effectiveness in its use of spectrogram image as compared with the FCNN model that uses Fourier transforms amplitude arrays even if the FCNN model takes less time to train compared with the 2D CNN model. Although it is worthy of note that the Neural network models applied are not fully optimized but rather there is room for improving the parameter and architecture of the model to increase the general accuracy achieved [8].

In a different approach, (Naidu, D. , 2022) has presented a voice analysis system that can be implemented using deep learning for the identification of vishing since instead of relying on phone numbers which are nowadays prone to spoofing vishing involves the actual content of the call [9].

Based on these developments (Manolaki-Sempagios, M. 2023) investigated fairness and bias in the models, the authors intend to build a new benchmark for audio deep-fake detection that should be more balanced in terms of the gender and accents of the speakers. A possible drawback is that it may not have good generalization capability across methods of speech synthesis other than the voice conversion models employed to synthesize the new test voices in the benchmark [10].

In terms of what is more relevant for human lives, (Groh, M. , Sankaranarayanan, A. , & Picard, R. , 2022) investigated the signal reality/digital fake socio-politics speeches of well-known politicians, qualified the distinguishability of political messages by multi-modality audio sources and base rates of mis-/disinformation streams. It is also evident that participants who the authors tested had a harder time identifying deep fakes to be fake, especially the text-to-speech deep fakes. The high base rate of deep fake has little interaction with the level of accuracy participants have in distinguishing between real and fake news stories. Introducing more modes (say audio, video) given to the participants increases their likelihood of detecting fake news [11].

3. The Research Problem

In today's digital world, new threats have appeared in the sphere of deepfake technology; specifically, the generation of AI fake voices threatening to become a real problem. This technology compromises the reliability of audio as a medium of passing information and is thus likely to be used in passing fake information, scams, and identity theft. That it is currently still difficult to identify or even verify the authenticity of an audio is an indication that we still have a glaring weakness in our arsenal. Of significant importance for research is filling this gap as it can help in the discovery of ways to prevent potential AI fake voices and thus, maintain the trust and security people have in their messages.

4. Proposed Solution

Our research work presents a new and improved deepfake audio detection model that employs deep learning methodology, particularly CNN on the spectral analysis of the audio signals. With the help of the rich ASVspoof 2019 database, our ultimate goal is to build a model that will be able to distinguish between human and synthetic speech with the involvement of TTS and VC techniques. The solution includes the spectrogram conversion of audio files, the CNN-based model applying Keras, data augmentation, and optimization strategies applied in this study. In doing so, this approach aims at building a stable and reliable system that is capable of identifying different types of synthetic speech, which will contribute to the improvement of voice recognition and authentication systems' security, help in fighting fake news, and further the cause of audio forensics.

5. Experiments

This section presents the experiments that involved the creation of and testing of the deepfake audio detection. We first describe the approach to collecting data, followed by the data preprocessing, architectures and applied methodology of the model, and the training methodology.

- **Data Collection and Preprocessing**

In developing the Deepfake audio detection system many prior researchers used the ASV spoof dataset of older versions. For our study, the ASV spoof 2019 dataset is used which is open source. It has various audio files which include real, (TTS) and (VC) audio data[6]. It is retrieved from the internet, saved in the hard disk directory, and then called from this place only. Because of the computational nature of the system, only a finite number of audio clips (31905 files) have been incorporated into this study along with the text file containing the description of each audio file. To this end, we used the text files to sort out the audio based on the system ID and then synthesized 3 folders. The audio files have not been standardized in the processing to retain the basic characteristic of the data as clips only contain a few seconds of audio. To correspond to every audio class, there is the creation of a specific folder with images containing the spectrograms of audio files of the particular class. And lastly, all the spectrograms are assembled in one folder to form the data frame of image data.

- **Architecture and Applied Methods**

After which with all the libraries imported, the generated spectrogram is used in the machine learning model. The third output-dependent classes in this problem are Real, Spoof-TTS, and Spoof-VC; this is a supervised multiclass classification problem. It is worth mentioning that the deep learning model is an extension and was developed over the Convolutional neural network. In this research, we constructed a model using the Keras framework. Keras on the other hand, is a cheap learning library that is not so tightly connected with the programming language Python but can work with TensorFlow. TensorFlow is the API constructed with the help of the AI system of Google and it has been built out of multiple numbers of dimensions which can also be considered bits of beams of one part of the data flow chart to the other part.

In this study, a sequential model is formed with four Conv2D layers with a respective number of layers of 32, 64, 128, and 256. Batch normalization layers are also included here and the required input values will be standardized here so the standardized values will be done automatically through a deep neural network. The new layer was then introduced into the network as the max pooling layer where contrary to convolutional layers the dimension of the feature maps of the input images was reduced but only contained the maximum spatial data. The layer models look like the below Figure 1:

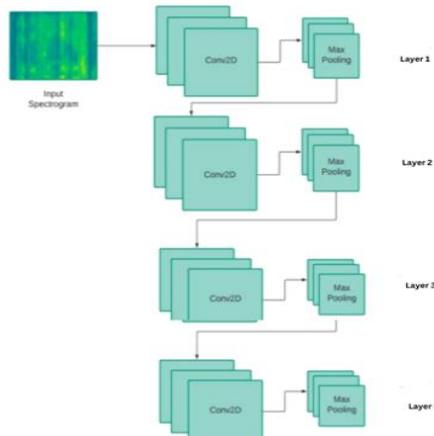


Figure 1 Keras CNN model layers

The function used for activation for conv2d layers was RELU and it gave a non-linear result wherein the model was compiled with categorical cross-entropy as the classes of the output are categorical and the number of labels is more than two and this is why binary cross entropy cannot be used here. Here the objective is to minimize the loss to achieve the optimal solution, the loss is cost and as highlighted, the variables corresponding to the real integers lost. That is called Adam here to get optimization Such loss function is used here In machine learning, the model summary provides the information on the total overall parameters which have been learned in the model. Data pre-processing of train and test datasets of 80:20 are done from data using the function `train_test_split` from the `sklearn` model selection. Again to perform this, the training data set was split into train; validation in the ratio of 90:10 where the latter was used in validating the model during its training. Keras has the image processing classes for the image data augmentation applied to the model for the image data generator class. This class is created in order such training, validation, and test data are available.

The model is trained with 50 epochs and because of so many epochs, the model overfits and to avoid overfitting, early stopping is applied which checks whether the model is enhancing further on a holdout validation set. For other cases the model was trained to epoch 38 the code then stated that it had stopped at that point. Learning rate is also used which depicts to what extent the model learned the problem with the help both the learning rate was passed as a hyperparameter and the early stopping was passed as the callbacks in the model. Using the forecasting function of the completed model, test datasets are utilized and it demonstrates how many data points were successfully forecasted in the three output variables.

- **Evaluation**

To ensure the accuracy of the model and verify the goodness of fit several methods used include the validation-loss, validation-accuracy, and the training-accuracy. It has to be noticed while training that particular model with the increasing number of epochs the validation loss should decrease and accuracy should increase. If we see that the validation set accuracy declines and loss rises then it becomes obvious that the given model does not accumulate any new information but just memorizes. On the other hand, in the case of overfitting, both would be on the rise since the nominal data and patterns would be overemphasized. As discussed earlier it can also be seen that the best situation concerning the model fitting is when the training loss and the validation loss are similar to each other. If the training loss is much more than the validation loss then it is a clear indication that the model is under-fitting but if the training loss is less than the validation loss then the model is overfitting. Things must be made perfect and that is why the great aim is to be ideal. The loss mentioned in the result section which is validation loss and accuracy is to validate the results.

6. Results and Discussion

From all the experiments that we conducted, we achieved a final training accuracy of 97% to 98%, and a training loss of 10% at 38 epochs after which the model did not improve any further. At the same time, the VL ranged from 20% to 30% while the accuracy was between 95 and 96% on the validation set. Shown below is the visual representation of how both the loss and the accuracy unfolded for the validation and training set per epoch.

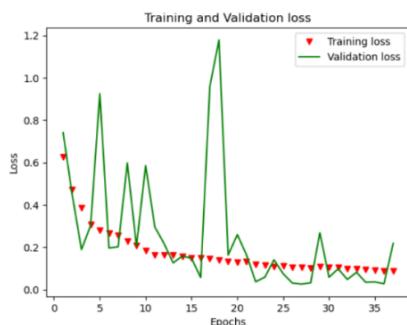


Figure 2 Training and Validation Loss

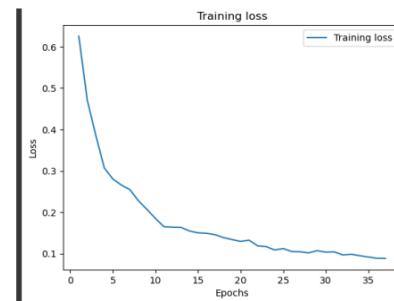


Figure 3 Training Loss

In Figure 3 there is the training validation loss as a function of epochs done in this experiment. This is the actual accuracy we need from our results, not some percentage increase we have to readjust for a different set of data. Here it is revealed that when we feed sample images to the model for validation then it gives 95%- 96% accuracy. We also observe the training accuracy of 97% to 98% which depicts how well the model trained on the training data.

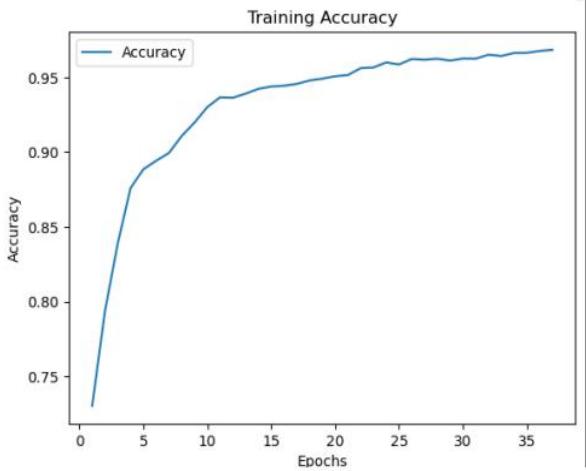


Figure 4 Training Accuracy

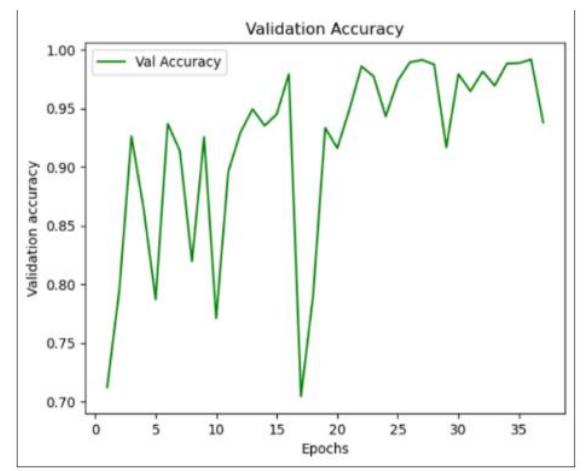


Figure 5 Validation Loss

As for the test data, the value of the given prediction accuracy comprises 99%. As it is evident from Figures 6 and 7 it can be seen that the proposed model has almost zero percent of misclassification.

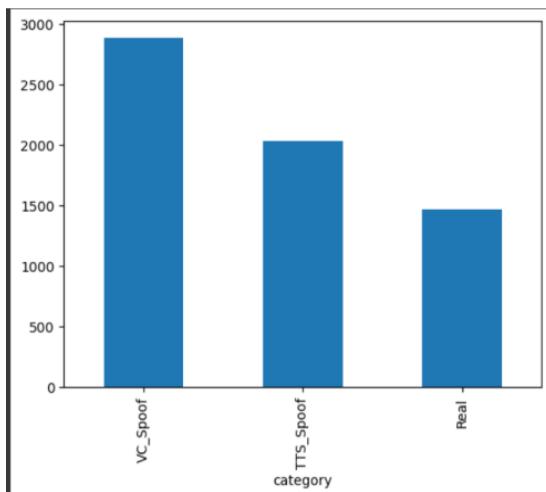


Figure 6 Actual Category

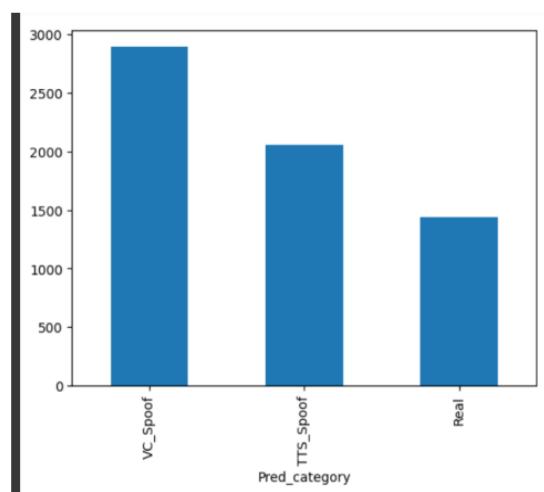


Figure 7 Predicted Category

7. Conclusions and Future Direction

This research was helpful in the process of demonstrating how a Convolutional Neural Network (CNN) model could be used to isolate fake and authentic audio in almost all of the cases. Taking advantage of the ASVspoof 2019 dataset, which took audio clips and converted them into spectrograms, the model that the study used for classification aimed at distinguishing real speech from TTS and VC. CNN model, with the assistance of the Keras platform, was used to detect deep fakes and it was found that the model has a training accuracy of 97- 98 % and test set accuracy of around 99 % to conclude the effectiveness of this model. These results indicate that it is viable to enhance audio security and steer clear of voice cloning misuses using modern methods of machine learning. In fact, CNN models can be fine-tuned to be a countermeasure to voicing cloning, especially in cases of forgery and identity theft. This could be very useful in safeguarding individuals, organizations, and governments against impersonation adversaries. In addition, the practice of the CNN models in audio forensics may also improve the investigations and the distinct approaches could provide better admissible evidence in the legal cases involving manipulation of audio, cybercrime, and counter-terrorism. These models require a number of improvements and deployment and to the advancement of audio authentication and cybersecurity.

The CNN-based model showed quite positive results in the case of demarcating between real and fake audio, which proves the effectiveness of using this kind of classification in enhancing audio security. In these cases, several future research directions are suggested for the advancement of this stream of research. If possible, to make the model more realistic when applied in the real world, the background noise removal feature will play a significant role. Analysed audio is inherently noisy and may contain low levels of noise in their real-world environment and can interfere with the functioning of the AVS. By using a noise reduction algorithm, it would be possible to have a higher reliability of the model for different settings thus enhancing its performance. To improve the model performance we need to incorporate more diverse audio samples in our dataset which comprise different languages, accents, and various background noise.

8. References

- [1] M. Dordevic, M. Milivojevic, and A. Gavrovska, "DeepFake Video Analysis using SIFT Features," 2019 27th Telecommunications Forum (TELFOR), 2019. doi: 10.1109/telfor48224.2019.8971206..
- [2] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," 2019 International Conference on Biometrics (ICB), 2019. doi: 10.1109/icb45273.2019.8987375.
- [3] S. Jadhav, R. Patole, and P. Rege, "Audio Splicing Detection using Convolutional Neural Network," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019. doi: 10.1109/icccnt45670.2019.8944345.
- [4] Z. Ali, M. Imran, and M. Alsulaiman, "An Automatic Digital Audio Authentication/Forensics System," IEEE Access, vol. 5, pp. 2994-3007, 2017. doi: 10.1109/access.2017.2672681.
T. Bhangale and R. Patole, "Tampering detection in digital audio recording based on Statistical Reverberation features," SpringerLink, https://link.springer.com/chapter/10.1007/978-981-13-3600-3_55.
- [5] J. Yamagishi, S. Todisco, H. Delgado, X. Wang, M. Sahidullah, T. Kinnunen, Z. Ling, and T. Kanervisto,
- [6] "ASVspoof 2019: The 3rd Automatic Speaker Verification Spoofing and Countermeasures Challenge database," Datasshare, University of Edinburgh, 2019. Available online: <https://datasshare.is.ed.ac.uk/handle/10283/3336>.
- [7] H. Dhamyal, A. Ali, I. A. Qazi, and A. A. Raza, "Fake audio detection in resource-constrained settings using microfeatures," ISCA Archive, https://www.isca-archive.org/interspeech_2021/dhamyal21_interspeech.html
- [8] T. Tami and L. T. Tijsmans, "Vocal fake detection using discrete Fourier transform in ...," <https://rp.os3.nl/2020-2021/p10/report.pdf>, <https://rp.os3.nl/2020-2021/p10/report>.
- [9] D. Naidu, "Voice analysis system for detection of vishing using Deep Learning," International journal of health sciences, <https://sciencescholar.us/journal/index.php/ijhs/article/view/7520>
- [10] M. Sempagios, "Benchmarking audio deepfake detection," Eindhoven University of Technology research portal, <https://research.tue.nl/en/studentTheses/benchmarking-audio-deepfake-detection>
- [11] M. Groh et al., "Human detection of political speech deepfakes across transcripts, audio, and video," arXiv.org, <https://arxiv.org/abs/2202.12883>.



FUSST International
Conference on Computer
Science and Engineering

FICSE

Research Article

Article Citation:
Afzaal et al. "Non-Invasive Glucose Monitoring Through IoT and Machine Learning"

Non-Invasive Glucose Monitoring Through IoT and Machine Learning

Fatima Afzaal ^{a,*}, Menahil Naseer ^a, Muhammad Kamran ^a

^a Department of Computer Science, Air University, Aerospace and Aviation Campus Kamra, Pakistan.
(fatimaafzaal2002@gmail.com).

* Corresponding author: fatimaafzaal2002@gmail.com

Abstract:

Diabetes is a condition that requires individuals to constantly monitor their health to avoid risks and complications. Many diabetic patients experience discomfort when checking their glucose levels through traditional invasive methods, such as finger-pricking, which can lead to infections and mental discomfort. Our research proposes the development of Gluco-Aid, a non-invasive glucometer that uses Near-Infrared (NIR) sensors to measure glucose (blood sugar) levels. Gluco-Aid consists of a device that monitors glucose levels in the body and reports the data to a mobile app integrated with an IoT platform. We will also train machine learning algorithms on a dataset of glucose measurements to improve accuracy and precision. The data will be used by the algorithms to detect patterns and enhance their glucose level prediction accuracy. This technology can monitor glucose levels in real time and, through machine learning, provide patients with alerts on when to eat based on the severity of health threats from high or low blood sugar readings.

Keywords: Non-Invasive Glucose Monitoring, Near-Infrared (NIR) Spectroscopy, Machine Learning Algorithms, IoT-Based Health Monitoring, Glucometer Technology, Diabetes Management Solutions, Innovative HealthCare, Digital Health Solutions

1. Introduction

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024 Afzaal et al.



Published by
Foundation University
Islamabad.

Web: <https://fui.edu.pk/>

Diabetes is a chronic disease that occurs due to the result of uncontrolled glucose levels in the body [1][2]. The prevalence of diabetes is projected rise to more than 7% of the adult population by 2030 [3]. Nearly all countries are grappling with serious complications such as heart diseases, kidney failure, blindness and amputation resulting from diabetes. In 2019 alone, diabetes and its complications caused a staggering 4.2 million adult lives [4]. Diabetes not only affects world's health but also causes huge financial loss. In 2015, the global cost of diabetes accounted for 1.8% of global GDP, estimated at about US\$1.31 trillion. Moreover, WHO estimated the cost of lost productivity for people with diabetes to be more than fivefold the direct cost of the disease [5][6].

Because the health and economic burdens for diabetes are so large, management of glucose must be instituted. Traditional glucose monitoring uses painful, inconvenient skin-invasive methods. Poor compliance causes patients to fail to follow the standard of care. The prevalence of diabetes is on the rise globally, and treatments that are more patient-friendly are in increasing demand. This has accelerated the innovation in non-invasive glucose meters that are continuous and in real time, with no blood sampling. They address the physical and practical obstacles posed on diabetic patients.

Unlike traditional invasive methods that rely on finger-pricking or continuous glucose monitors (CGMs), which require sensors to be placed under the skin, our system offers a pain-free alternative. The use of Near-Infrared spectroscopy removes the risk of infection and discomfort, making glucose monitoring more accessible and user-friendly for patients. Non-invasive glucose monitoring (especially by means of Near-Infrared [NIR] spectroscopy) has attracted a great deal of attention for providing measurement capabilities beyond the skin surface. There are several methods employed for non-invasive glucose monitoring, but NIR spectroscopy is gaining increasing popularity due to its promising accuracy and non-invasive nature, making it a preferred choice among both researchers and patients. NIR spectroscopy is used to study absorption of light at specific wavelengths and due to its non-invasive nature, it provides a rapid and pain-free alternative over the traditional methods.

By combining machine learning with NIR sensors, we can process sophisticated data resulting in reliable glucose readings and dynamic feedback adapted to individuals. The incorporation of machine learning algorithms allows our system to continuously improve prediction accuracy by learning from the growing dataset of glucose measurements.

Unlike static systems, this approach offers dynamic adaptability, ensuring high accuracy across various skin types, environmental conditions, and patient profiles.

Our objective is to design a non-invasive glucometer using Near Infrared Reflection (NIR) sensors, integrated with machine learning algorithms, to enable precise real-time glucose monitoring. It aims to promote better management of diabetes through less cumbersome and more accessible means of monitoring glucose, therefore contributing to improved patient outcomes.

The research, therefore, contributes to continuously growing efforts towards the application of artificial intelligence and machine learning in healthcare, with particular emphasis on developing different solutions for non-invasive glucose monitoring. Going beyond an improvement in the accuracy of glucose measurement, the proposed system would be able to make personalized recommendations, enhancing overall management of diabetes. Future work will explore further advancements in this field to continue improving the quality of life for diabetic patients. This research develops a non-invasive glucose-monitoring system integrating AI with health-monitoring devices to improve diabetes management. The key objectives include creating an AI algorithm for real-time glucose prediction, developing a personalized mobile app, and ensuring system reliability. The study contributes both conceptually and empirically, offering advancements in AI-driven glucose monitoring and user-specific healthcare solutions.

2. Related Work

This section discusses the existing and traditional methods for monitoring blood glucose levels characterized as methods that require penetration of the skin and methods that do not.

Traditional blood sugar monitoring methods are typically invasive, with the most common being finger-pricking to draw blood and placing a blood drop on a strip that is inserted into a glucose meter. However, this method can lead to infections, requires a large number of test strips over time, and the cost of these strips can become prohibitive for patients [7][8][9].

Another invasive method is continuous glucose monitoring (CGM), which works through a small sensor that measures the blood glucose level. This method offers information on blood glucose, including real-time readings of glucose levels and fluctuations in blood sugar. It also includes alerts, and alarms. Its limitations include the cost of the sensors as well as skin irritation as some patients may develop allergies to the patches used to attach CGM sensors which causes discomfort, itching, redness and rashes in some cases [10][11].

M Tierney et al. proposed the GlucoWatch biographer, a glucose monitoring system that does not require skin penetration and uses reverse iontophoresis to draw glucose through the skin, which is then detected by an amperometric biosensor. It can provide glucose readings every 20 minutes for up to 12 hours. The study demonstrated that the GlucoWatch biographer achieved high accuracy and precision in both clinical and home environments. However, device accuracy was affected at low glucose levels, as it showed a slight positive bias when the glucose level decreased to 10 mmol. Additionally, the study reported skin irritation from using the device [12].

Xi Peng et al. proposed the use of fiber lasers for monitoring blood glucose without skin penetration. According to research, fiber lasers have great importance in the medical field because they provide a high-quality light source compared to other light sources and also have low cost and excellent performance. Fiber lasers have been applied to various optical methodologies for glucose monitoring, including Raman spectroscopy, photoacoustic spectroscopy, mid-infrared absorption spectroscopy, near-infrared absorption spectroscopy, and fluorescence spectroscopy. Such techniques benefit from the use of fiber lasers in yielding better and less invasive measurements. However, challenges such as variations in skin pigmentation and thickness can affect measurement accuracy, and the higher cost of fiber lasers compared to other light sources may limit their adoption [13].

Anmole S. Bolla et al. discussed that mid-infrared spectroscopy is a technique for measuring glucose concentration in blood. Basically, MIR spectroscopy measures infrared radiation absorbed by the molecular vibrations, and glucose possesses particular vibrational modes detectable in the MIR range. By analyzing the position and intensity of the absorption peaks in the MIR spectrum, the glucose concentration can be measured very accurately. MIR spectroscopy is a non-invasive technique for the measurement of glucose [14].

Quoc-Hung Phan et al. discussed that Glucose molecules have a special chiral structure that rotates the plane of polarized light, so measuring this rotation can tell us about the glucose readings. Polarimetry is a way to measure how light rotates when it passes through something. The linear polarization of light is analyzed in relation to path characteristics, temperature, and concentration. Researchers are exploring the use of polarimetry for measuring glucose levels without invasive procedures. However, the skin does not only consist of glucose, so it's hard to use polarimetry to only measure glucose. Scientists are working on ways to figure out how to tell the glucose readings without the interference of any other factor. If they can figure it out, this could be a really easy and painless way to measure glucose for people with diabetes [15].

Andreas Caduff et al. discuss the development of a multisensor system for non-invasive glucose monitoring, specifically designed for diabetic patients with Type 1 diabetes. This system incorporates two types of sensors: dielectric and optical.

To analyze the signals from these sensors, a multiple linear regression analysis model is employed. One limitation of this work is the small study group, consisting of only ten male patients with Type 1 diabetes, which may affect the overall accuracy of the system [16].

3. The Research Problem

Diabetes management is a crucial part of healthcare, as it involves continuously monitoring blood glucose levels to prevent complications. Traditional methods like the finger-prick test and continuous glucose monitors (CGMs) have been widely used, but they come with significant drawbacks. These are invasive techniques, for the most part painful, which can be a deterrent to a patient having routine blood glucose tests. The cost and hassle on the part of patients who use such methods add to their discomfort.

Discomfort and inconvenience associated with traditional methods, patients monitor their sugar level irregularly, which may result in poor diabetes management. This underlines the critical need within diabetes care for a reliable, non-invasive means of monitoring glucose levels that is both comfortable and usable consistently by the patients.

Although several non-invasive methods are being tested, such as optical, electromagnetic, and bioimpedance techniques, none of those techniques have gained wide acceptance so far in the clinical environment. The main obstacles have been, so far, precision and reliability issues, with the ability to provide continuous real-time data. However, NIR spectroscopy is a promising technique since it is non-invasive and can offer continuous glucose monitoring while analyzing light interaction with body tissues.

The main problem this research would try to solve is the creation of a non-invasive glucometer that effectively obtains glucose level measures by using NIR spectroscopy together with advanced machine learning algorithms in an effort to enhance the accuracy level of non-invasive methods and ensure the feasibility of the device for everyday application by diabetic patients.

This research is crucial because it addresses a significant unmet need in diabetes care: the development of a non-invasive, user-friendly glucose monitoring system that can be easily integrated into daily routines. Such a system can make compliance much easier, while the resulting overall health outcomes can be greatly improved by avoiding the pain and nuisance of traditional methods. In a larger perspective, such successful development will catalyze further nonlinear advances in the area of non-invasive health monitoring for the benefit not only of diabetic patients but also for the digital health sector as a whole.

4. Proposed Solution

This study proposes a non-invasive glucose monitoring solution using Near-Infrared (NIR) spectroscopy and machine learning to enhance glucose level predictions. The system captures glucose absorption patterns via NIR sensors and combines this data with physiological factors like age, weight, and heart rate. To address noise from ambient light and skin variations, preprocessing steps such as noise reduction and normalization were applied. Principal Component Analysis (PCA) was used to extract the most relevant features for glucose prediction.

Machine learning models, including Random Forest, Linear Regression, Gradient Boosting, and Ridge Regression, were trained on the preprocessed data, with the dataset split into 70% training and 30% testing. Five-fold cross-validation ensured the models generalized well. To enhance performance, a Voting Regressor was used to combine model predictions based on their cross-validation performance, improving accuracy and robustness.

Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2), with the Voting Regressor outperforming individual models. This methodology provides significant advantages over traditional glucose monitoring by offering non-invasive, real-time data collection and continuous monitoring. The integration of machine learning allows for model updates, improving accuracy over time. Future work will focus on expanding the dataset, optimizing models, and integrating with digital health platforms.

Multiple solutions have been attempted to achieve better performance of non-invasive glucose monitoring. Again, the integration of NIR with other non-invasive techniques, such as bioimpedance, could offer greater reliability and fewer errors, adding complexity in integrating data and device design. Other advanced machine learning methods, such as ensemble learning, were also promising, showing higher accuracy but at the cost of heavy computational resources and large datasets. In selecting the optimal solution, accuracy, non-invasiveness, practicality, and scalability have been considered in this proposal. The proposed solution, which couples near-infrared spectroscopy with machine learning algorithms, outperforms other models by making a compromise between model performance and comfort for users in practice, therefore becoming a good candidate for wide clinical application.

It entails a number of advantages, such as less discomfort compared with its alternative methods, constant monitoring, and in real-time, the data medical experts require on diabetes management. The integration of machine learning ensures high accuracy for a range of conditions and will go further in enhancing the device's effectiveness and reliability.

It is a breakthrough in the area of non-invasive glucose monitoring and has the potential to revolutionize diabetes management with more comfort and precision than invasive blood testing.

5. Results and Discussion

The study on a non-invasive glucometer using Near-Infrared (NIR) technology and machine learning showed encouraging results. The NIR sensor quantified glucose by measuring the changes in the reflected infrared light through the skin and provided very similar results to those obtained using traditional glucometers in a variety of environmental situations, such as skin type and temperature.

A series of experiments was conducted in order to test the efficacy and performance of our system. This involved comparing glucose measurements obtained using our system with those obtained using standard glucometers in real-life conditions. The diabetic patients tried the Gluco-Aid system (volunteer), and shared the data related to its usability and effectiveness. Moreover, the experiments included tests concerning the responsiveness and accuracy of the system compared to traditional methods.

In practical terms, the Gluco-Aid system could be integrated into healthcare systems, offering continuous real-time glucose monitoring that allows for more dynamic and personalized diabetes management. For instance, the device could send real-time data to healthcare providers or caregivers, enabling timely interventions when glucose levels reach critical thresholds. This could reduce the need for frequent hospital visits and improve overall disease management for diabetic patients.

Machine learning algorithms were pursued for the prediction of glucose levels from NIR sensor data. These algorithms are trained using large datasets, which are then validated against conventional readings. Figure 2 gives a comparison of our system with traditional glucometers; this shows an error margin of ± 20 values.

Notwithstanding the promising results, certain limitations in the study were identified. The sensors were sensitive to ambient light and skin pigmentation, which required a lot of data for the models' training process. Future improvements should target improving sensor robustness and increasing model generalization across various populations.

Performance evaluation included assessing the ensemble learning model with metrics such as R^2 , Mean Squared Error (MSE), and Mean Absolute Error (MAE). These metrics confirmed the system's accuracy and reliability, demonstrating its potential as a non-invasive alternative to traditional glucose monitoring methods.

Statistical Comparison of Model Predictions with Actual Glucose Levels

i. Mean Comparison:

Gluco-Aid vs. traditional glucometer: comparing the means of actual blood glucose concentration from traditional glucometer and predicting using our system

$$\text{Mean} = \frac{1}{N} \sum_{i=1}^N x_i$$

Where x_i represents the actual glucose level and N represents total values. For this study:

- Traditional Glucometer Readings Mean = 162.75
- Gluco-Aid Readings Mean = 166.86

The proximity of these means suggests that, on average the system's prediction is aligned with traditional glucometer readings.

ii. Variance and Standard Deviation:

To quantify the variability between real and predicted glucose levels, we can use variance and standard deviation. Here's how we can compute them:

$$\text{Variance} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$\text{Standard Deviation} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

For this study:

- Traditional Glucometer Variance = 7515.87
- Gluco-Aid Variance = 7455.78
- Traditional Glucometer Standard Deviation = 86.69

- Gluco-Aid Standard Deviation = 86.35

The variance and standard deviation values for both the traditional glucometer and Gluco-Aid are quite similar. This joint denomination suggests that both systems fit the variability in glucose levels well. The closeness in these statistical measures indicates that Gluco-Aid performs comparably to the traditional glucometer in terms of capturing the variability in glucose readings.

iii. Error Metrics:

The error metrics are used to check the precision of a model:

Mean Squared Error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - x_{i'})^2$$

For this study:

$$MSE = 72.57$$

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_{i'})^2}$$

For this study:

$$RMSE = 8.52$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - x_{i'}|$$

For this study:

$$MAE = 6.32$$

Lower scores for MSE, RMSE, and MAE indicate that the system's predictions are closer to the actual glucose levels on average, reflecting the accuracy of our evaluation metric. Therefore, we aim for these metrics to be low, highlighting that RMSE and MAE are crucial for assessing prediction performance.

Also Figure 1 illustrates a comparative analysis of glucose level predictions made by the traditional glucometer and the Gluco-Aid system. This close alignment in the statistical measures highlights the reliability of the Gluco-Aid system in terms of both central tendency (mean) and variability (variance and standard deviation), demonstrating its capability to predict glucose levels in a manner consistent with the traditional glucometer.

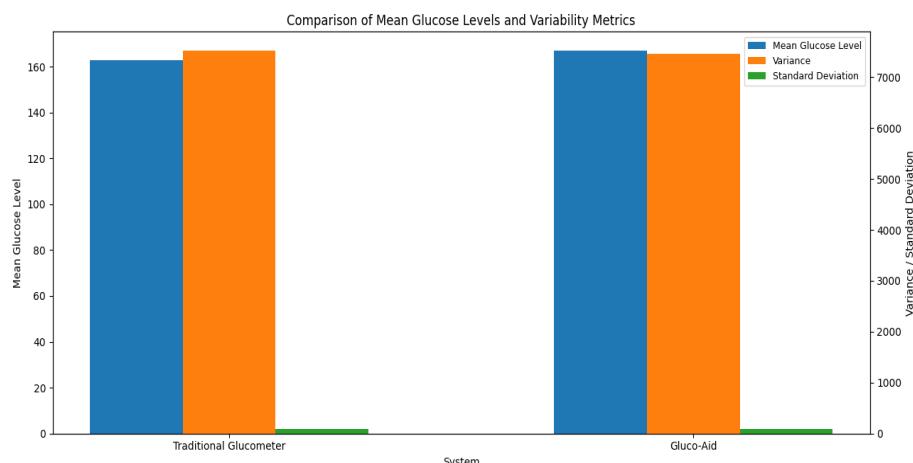


Figure 1 Comparison of Gluco-Aid with traditional system

Justification and Conclusion:

- **Performance is Good:** It signifies that the system has predicted glucose with decent accuracy and consistency, given that the means are close, variance values are similar, and error metrics are low.
- **A Room for Improvement:** Although the differences in mean values and errors are minor, suggesting strong performance, there is potential for further improvement. By adding in more accurate and robust model features, this could easily be achieved.
- **Findings:** The statistical comparison substantiates that the performance of the Gluco-Aid system is akin to that of traditional methods. This high level of certainty regarding the system's effectiveness and reliability emphasizes its potential as a feasible non-invasive solution for glucose monitoring. Although the results are compelling, continual refinement will contribute to enhancing the system's performance and applicability.

In summary, the Gluco-Aid system closely parallels conventional glucose measurement techniques, presenting a non-invasive option that demonstrates promising accuracy. Ongoing development and refinement are crucial for enhancing its efficacy and ensuring user comfort.

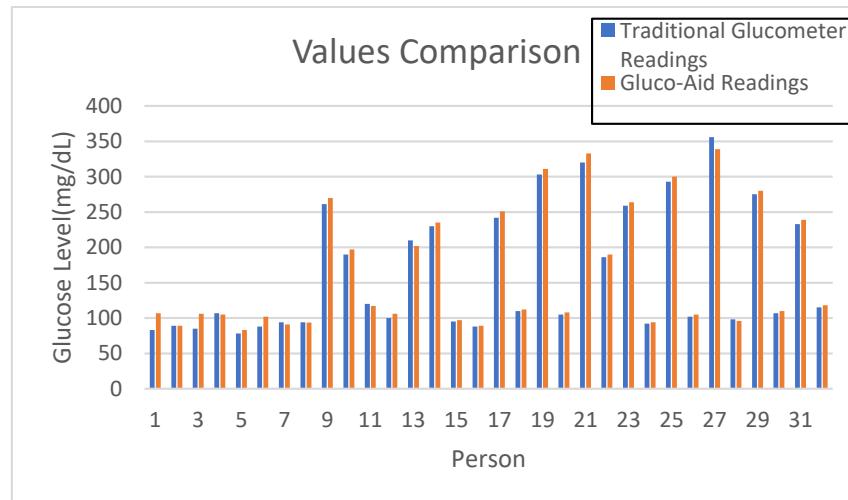


Figure 2 Testing Results Comparison

6. Conclusions and Future Direction

This study focused on developing a non-invasive but efficient blood glucose measurement system by incorporating Near-Infrared spectroscopy with machine learning algorithms. This may lead to a great finding in the advancement of the management of diabetes, considering that this modality is much less painful and more convenient than conventional methods of monitoring glucose. Our findings suggest that such a system may provide reliable glucose readings, which might improve patient compliance and enhance overall quality of life. This system not only removes the discomfort and risk of traditional invasive glucose monitoring methods but also overcomes the limitations of other non-invasive technologies. By integrating NIR spectroscopy with machine learning, our approach provides more accurate and personalized glucose readings, adapting to individual patient conditions and environmental variations.

Future Work:

1. **Improving Accuracy:** Future research will pin down NIR assessment accuracy, especially in optimizing spectroscopy and machine learning. It shall comprise calibration and validation for consistent glucose measurement.
2. **Integration with Digital Health Platforms:** We plan to explore the integration of this system with digital health platforms to enable seamless data sharing and real-time monitoring. This could facilitate better management and tracking of diabetes, leading to more personalized care.
3. **Multi-Technique Approaches:** NIR spectroscopy could be used together with other non-invasive techniques such as bioimpedance or optical methods to improve the reliability of glucose monitoring. In this method, the accuracy would increase and the error margins would fall.
4. **Technological Advances:** Continuous technological development will further expand this non-invasive health application area. Innovations will give birth to new personalized health management solutions.

In summary, this study lay the foundation for a paradigm shift in glucose monitoring, with great potential impact on the management of diabetes in the future. Further development through precision, integration, and development of technology is needed to tap the full potential of this non-invasive system and extend its benefits to a variety of health-related uses.

7. References

- [1] A. Keshet, S. Shilo, A. Godneva, Y. Talmor-Barkan, Y. Aviv, E. Segal, and H. Rossman, "CGMap: Characterizing continuous glucose monitor data in thousands of non-diabetic individuals," *Cell Metabolism*., vol. 35, no. 5, pp. 758-769, 2023.
- [2] Y. Rostam-Abadi, J. Gholami, M. M. Jobehdar, M. Ardeshir, A. M. Aghaei, S. Olamazadeh, et al., "Drug use, drug use disorders, and treatment services in the Eastern Mediterranean region: a systematic review," *The Lancet Psychiatry*, vol. 10, no. 4, pp. 282-295, 2023.
- [3] Y. Rostam-Abadi, J. Gholami, M. M. Jobehdar, M. Ardeshir, A. M. Aghaei, S. Olamazadeh, et al., "Drug use, drug use disorders, and treatment services in the Eastern Mediterranean region: a systematic review," *The Lancet Psychiatry*, vol. 10, no. 4, pp. 282-295, 2023.
- [4] Wei, J. Tian, C. Tang, X. Fang, R. Miao, H. Wu, et al., "The influence of different types of diabetes on vascular complications," *Journal of Diabetes Research*, vol. 2022, no. 1, p. 3448618, 2022.
- [5] C. Bommer, E. Heesemann, V. Sagalova, J. Manne-Goehler, R. Atun, T. Bärnighausen, and S. Vollmer, "The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study," *The Lancet Diabetes & Endocrinology*, vol. 5, no. 6, pp. 423-430, 2017.
- [6] W. Q. Lin, Z. J. Cai, T. Chen, M. B. Liu, N. Li, and B. Zheng, "Cost-effectiveness of dipeptidylpeptidase-4 inhibitors added to metformin in patients with type 2 diabetes in China," *Frontiers in Endocrinology*, vol. 12, p. 684960, 2021.
- [7] Sunny, S., & Kumar, S. S. "Optical based non invasive glucometer with IoT". *International Conference on Power, Signals, Control and Computation (EPSCICON)* (pp. 1-3). Jan 2018 IEEE.
- [8] X. Ding, B. W. -K. Ling and C. Huang, "Non-invasive Blood Glucose Estimation Using Statistical Features Defined via Convex Combination of One Norm and Infinity Norm Optimization Problems," *2022 IEEE International Symposium on Product Compliance Engineering - Asia (ISPCE-ASIA), Guangzhou, Guangdong Province, China, 2022*, pp. 1-4, doi: 10.1109/ISPCE-ASIA57917.2022.9970814.
- [9] MM. Aihara, T. Irie, K. Yasukawa, I. Minoura, N. Miyauchi, M. Nishi, N. Katayama, et al., "Development of a high-performance liquid chromatographic glycated albumin assay using finger-prick blood samples," *Clinica Chimica Acta*, vol. 542, p. 117272, 2023.
- [10] P. Jain, A. M. Joshi, and S. P. Mohanty, "IGLU 1.0: An accurate non-invasive near-infrared dual short wavelengths spectroscopy based glucometer for smart healthcare," 2019, arXiv:1911.04471.
- [11] D. Rodbard, "Continuous Glucose Monitoring: A Review of Successes, Challenges, and Opportunities," *Diabetes technology and therapeutic*, vol. 18, no. S2, pp. S23–S213, 2016, doi: 10.1089/dia.2015.0417.
- [12] M. J. Tierney, J. A. Tamada, R. O. Potts, L. Jovanovic, and S. Garg, "Clinical evaluation of the GlucoWatch® biographer: A continual, non-invasive glucose monitor for patients with diabetes," *Biosensors and Bioelectronics*, vol. 16, no. 9–12, pp. 621–629, 2001, doi: 10.1016/S0956-5663(01)00189-0.
- [13] X. Peng, Y. X. Yan, and H. Liu, "On the use of fiber lasers in non-invasive blood glucose monitoring," *Optical Fiber Technology*, vol. 68, p. 102822, 2022, doi: 10.1016/j.yofte.2022.102822.
- [14] A. S. Bolla and R. Priefer, "Blood glucose monitoring- an overview of current and future non-invasive devices," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 5, pp. 739–751, 2020, doi: 10.1016/j.dsx.2020.05.016.
- [15] Q. H. Phan and Y. L. Lo, "Stokes-Mueller matrix polarimetry system for glucose sensing," *Optics and Lasers in Engineering*, vol. 92, pp. 120–128, 2017, doi: 10.1016/j.optlaseng.2016.08.017.
- [16] A. Caduff *et al.*, "Non-invasive glucose monitoring in patients with Type 1 diabetes: A Multisensor system combining sensors for dielectric and optical characterisation of skin," *Biosensors and Bioelectronics*., vol. 24, no. 9, pp. 2778–2784, 2009, doi: 10.1016/j.bios.2009.02.001.

Reconstructing Facebook Artifacts: A Volatile Memory Forensics for Cybercrime Investigation on Windows

Muhammad Ahsan Qureshi ^{a*}, Najmul Hassan ^b, Kashif Nasr ^c, Adeel Ahmed,
Mehmood-ul-Hassan ^d

^a Department of Computer Science, Air University Islamabad, Pakistan.

^b Sunway University, Malaysia.

^c COMSATS University Islamabad, Pakistan.

^d National Response Center for Cybercrimes (NR3C), FIA, Pakistan

* Corresponding author: ahsan.qureshi@aack.au.edu.pk

Abstract:

Digital evidence from volatile memory is crucial in cybercrime investigations. Understanding and reconstructing artifacts left in computer's RAM after using Facebook via an Internet browser is a significant aspect of this field. This research emphasizes the importance of interpreting these artifacts as a key source of evidence. Collecting and preserving forensically sound digital evidence from a running computer presents significant challenges. This study followed the standard digital forensics process, performing various Facebook activities such as login, wall posts, comments etc. The results demonstrate how Facebook artifacts can be identified using the AccessData FTK keyword search function. The primary objective is to analyze digital artifacts from the computer's RAM, classify them according to user actions on Facebook, and attribute the collected evidence to a suspect. This evidence can then be used in legal proceedings.

Keywords: Facebook; RAM; cybercrime; digital forensics; information security; privacy.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright

Copyright © 2024 Qureshi et al.



Published by
Foundation University
Islamabad.

Web: <https://fui.edu.pk/>

way that can bring change in their behavior. The increase in traditional crimes that have gone online with the advancement of technology has made it necessary to develop a new method of conducting forensic investigations. Digital traces on computers are often the primary basis for digital forensic investigations, and they are useful in reconstructing a digital crime scene. The volatile nature of RAM renders it as one of the most important yet difficult media for digital evidence. Unlike data stored in non-volatile storage devices like hard disks and USB drives, RAM data is continuously changing and can be lost due to power outages. This research focuses on Facebook artifacts analysis from a Windows 10 OS using Google Chrome Internet browser and how computer stores the Facebook

artifacts in RAM, and how these artifacts can be interpreted to reconstruct Facebook events during cybercrime investigations.

2. Related Work

Ivan et al. emphasized the significance of RAM in forensic investigations, highlighting its role in storing crucial data like running processes, network connections, and user activities. The study discussed various methods and tools for analyzing RAM data, though it noted a lack of comprehensive evaluation of these tools' reliability, suggesting the need for future research to enhance live forensics [4]. Hitesh Sanghvi et al. explored the challenges of forensic analysis in Mozilla Firefox's private browsing mode. Their experiments revealed that even in private mode, forensic tools like FTK could retrieve substantial evidence from RAM and hard disks, demonstrating the potential for recovering data across various online activities, including Facebook [5]. Kelvin et al. focused on Facebook forensics, identifying the types of artifacts that can be retrieved from RAM dumps and Internet browsers, such as JSON and HTML formats. Their study demonstrated the recovery of Facebook activity evidence like posts, comments, and messages [6]. Ranul et al. discussed the challenges of extracting evidence from volatile memory in social media applications, proposing a framework to aid in the collection of admissible evidence from RAM without data loss, ensuring its integrity for legal proceedings [7]. Neethu et al. presented methodologies for extracting user credentials and other Internet-related evidence from Windows system memory dumps. The study highlighted the recovery of critical information such as visited sites, emails, search queries, and Facebook fragments, emphasizing the reliability of pattern-based searching in cyber forensics [8]. Huwida Said et al. investigated forensic analysis of Internet browsing in private modes across Internet Explorer InPrivate, Google Chrome Incognito, and Mozilla Firefox Private Browsing. Their study, conducted on Windows XP VMs, focused on traces like visited URLs, cached web pages, and search keywords. They found that while no traces were present in web history or cache, some were detected in the pagefile.sys and physical memory [9]. Hai-Cheng Chu et al. examined live data acquisition from RAM to reconstruct previous Facebook sessions, providing valuable insights for digital forensics in cybercrime investigations. Their research highlighted the importance of specific keywords in tracking Facebook activity [10]. Noora Al-Mutawa et al. explored extracting instant messaging artifacts from hard disks, specifically focusing on Facebook Chat. They found that while artifacts were scattered across the hard disk, full chat scripts were typically located in specific areas, and noted the challenges in identifying non-Latin character sets like Arabic [11]. James et al. introduced a signature-based approach to reconstruct user events from low-level traces, showing potential for detecting user actions from post-mortem forensic analysis, though further development is needed [12]. Ming Sang Chang et al. investigated Twitter forensics on Windows 10 using Internet Explorer and Google Chrome. They found that Twitter activity left valuable evidence on hard drives and memory dumps, useful for profiling user behavior, with varying implementations across devices [13]. Teing Yee Yang et al. analyzed artifacts from Facebook and Skype on Windows 8.1, revealing types of data remnants such as installation logs, contact lists, and message content, which can be critical for investigations [14]. Anderson et al. reviewed authorship attribution methods in social media forensics, highlighting the need for advanced algorithms that can handle small sample sizes and diverse data types to identify authors in anonymous online environments [15]. Martin et al. proposed a hybrid system for harvesting data from social networking sites, combining a custom add-on with web crawling. Their method effectively collected profile data and metadata, demonstrating improvements over traditional web scraping techniques [16].

3. The Research Problem

It is very crucial to have the identification and interpretation of Facebook digital evidence in a running computer as an area of study in forensic science. The traditional approach to forensic practices has been retrieving data from offline systems while, thus not having any knowledge on how to interpret and analyze the digital artifacts from active systems. This thus provoked the challenge on attributing evidence to suspect for legal purposes. The main objective of this research study is to identify types of Facebook artifacts that can be collected from RAM on Windows based systems and coming-up with methods that will help pinpoint these pieces of evidence to the exact actions or experiences that could be attributed to individual suspects involved in such activities. Therefore, the ability to reliably interprets the artifacts left by Facebook can help enhance digital evidence presented in court, thus allowing more effective prosecution of cybercrimes.

4. Proposed Solution

This proposed key steps and research methodology of forensic analysis of RAM is shown in Figure 1.

a). Facebook Activities and RAM Acquisition:

The first thing is to determine the types of Facebook artifacts that get stored in RAM during user activities such as logging in, posting and commenting, etc. This provides an in-depth analysis of how Facebook interacts with RAM and leave traces of user activity. During runtime, acquiring a snapshot of RAM specialized forensic tool will be applied, preserving its content for further analysis.

b). Data Reconstruction and Analysis:

This research is aimed at reconstructing the Facebook events analyzed in the retrieved RAM artifacts after the acquisition of the RAM data. The interpretation and correlation of the RAM artifacts shall be performed manually for specific user actions. Keyword searches through strings and memory forensic tools shall be used to locate and interpret relevant artifacts.

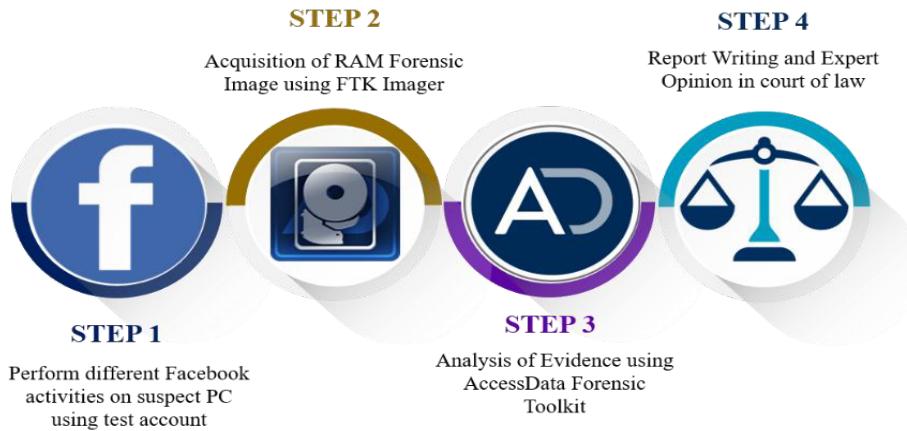


Figure 1: Research Methodology

5. Experiments

The testing in this study is conducted on realistic data and methodology, with the process strictly adhering to the standard process of a digital forensic investigation. A variety of Facebook activities were performed and analyzed to understand the format of the artifacts left in the RAM. The research methodology is divided into two phases i.e. memory image acquisition and memory analysis of the acquired RAM image. The memory acquisition phase basically dumps the RAM content, along with the pagefile.sys file, using the FTK Imager tool. During the phase of memory analysis, a licensed version of AccessData Forensic Toolkit is used to analyze Windows 10 volatile memory. While extraction of pagefile.sys file is a very important evidence source that cannot be overlooked in this task. It is best practice to recover deleted data using data carving or other techniques to achieve an optimum level before starting forensic analysis.

5.1 Experimental Setup

This research has been carried out on a personal computer virtual machine. The purpose of performing the experimentation in a virtual machine is that it is resource-efficient, as it functions like a real and complete computer system. Table 1 shows the experimental setup and system specifications used to conduct this research.

S.No.	Tools/Devices	Specifications
1.	Hypervisor	VMware Workstation Pro 16.0
2.	Hardware	Desktop-HP Omen 15
3.	Operating System	MS. Windows 10 Home, version 22H2
4.	System Type	x64-based processor
5.	RAM Size	4.0 GB
6.	Hard Disk (NTFS)	70.0 GB
7.	Google Chrome	v126.0.6478.115 (Official Build)
8.	RAM Image Acquisition Tool	AccessData FTK Imager v4.7.1
9.	Forensics Analysis Tool	AccessData FTK v6.0.0.52
10.	Social Media	Facebook Website (www.facebook.com)

Table 1: Experimental setup

Across the globe, AccessData FTK and EnCase are considered benchmark forensic tools as they can fulfill all digital forensic requirements regarding legal admissibility. Both produce relevant, reliable, and court-admissible reports of international standards of evidence handling and presentation. This ensures that digital artifacts extracted during investigations, including volatile memory, stay intact throughout the legal process. The proved credibility of FTK and EnCase in forensic investigations, including also robust data recovery and analysis capacities, makes them trusted leaders in cybercrime investigations and legal proceedings around the world. AccessData FTK was chosen for this study because of its robustness, competent functionality, and thorough capability for memory analysis. It has efficient data carving, indexing, and filtering functions that allow for the precise extraction and analysis of Facebook artifacts. Further, FTK supports multiple file formats, and with a user interface, big datasets are handled efficiently, whereas EnCase or Volatility cannot match it in terms of compatibility and setup complexity, especially when the focus is on RAM analysis.

5.2 Facebook Activities

We created two fictional Facebook users and a set of unique keyword strings that would be used as the Facebook post, and comments. An extensive keyword search was performed using both the index and live search features of AccessData FTK on the entire acquired RAM image of the suspect computer, resulting in an average of many hits for each keyword. Table 2 identifies the experimental activities performed on Facebook for forensic analysis.

Table 2: Experimental Facebook Activities

S.No.	Facebook Activity	Objectives
1.	Login	to analyze the artifacts of login attempts
2.	Profile	to analyze the artifacts of facebook profile
3.	Post & Comments	to analyze the artifacts of post & comments

The scope of investigations and keywords is always defined to search for related artifacts, however, the search should not be limited to these defined keywords. The lists of defined unique string set we used are shown in Table 3. Each string was used individually in a different activity.

Table 3: Unique Keywords for Facebook Activities

S.No.	Artifact Description	Keywords
1.	Login	facebook, https://www.facebook.com
2.	Username & Email ID	T Ali, testali334@hotmail.com
3.	Text Post	Hi, this is Test, How are you
4.	Comment on Text Post	I am fine, pleasure
5.	Post & Comments	to analyze the artifacts of post & comments

5.3 Memory Acquisition

Memory acquisition was conducted under controlled conditions to ensure the accurate retrieval of artifacts on Facebook. The experiments were performed on a Windows 10 machine using Google Chrome as the browser while focused on specific default settings. No extensions were installed in the browser so as not to interfere with artifact collection because certain extensions might enhance storage for corresponding data in RAM. In the context of simulating natural-user behavior, all the browser settings were still left to default cache, cookies, and history. This setting assures that the conditions adopted during the experiment relate to the real and honest operation of real systems in cybercrime investigations. It is logical to assume that different browser settings may have same type of artifacts but the path of the artifacts may change with respect to the browsers.

6. Results and Discussion

This section discusses the results of our experiment for each set of keywords. We describe the artifacts of the Facebook extracted from the RAM. We also discuss the interpretation of Facebook artifacts. The following information could be obtained from the data remnant in the computer's RAM after accessing Facebook from the Google Chrome Internet browser.

6.1 Facebook Login Artifacts

The artifact data shown in Figure 2 below reconstructs the complete event of logging into Facebook. The suspect first searched for Facebook on the Google search engine using the Google Chrome browser, then accessed Facebook. The artifact "<https://www.facebook.com?sk=welcome>" indicates that the user has successfully logged into Facebook.

File Content	
	Hex Text Filtered Natural
27c4b360	3F 00 25 00 20 00 42 00 6C 00 69 00 6E 00 6B 00 20 00 73 00 65 00 72 00 69 00 61 00 6C 00 69 00 7A 00 65 00 64 00 20 00 ?%··B-link··s-e·ri·al-i·z-e·d··
27c4b388	66 00 6F 00 72 00 6D 00 20 00 73 00 74 00 61 00 74 00 65 00 20 00 76 00 65 00 72 00 73 00 69 00 6F 00 6E 00 20 00 39 00 f-o·rm··s-t-a-t-e··v-e·rs-i·o-n··9·
27c4b3b0	20 00 0A 00 0D 00 3D 00 26 00 00 00 10 00 00 00 00 00 00 00 00 00 00 6C 00 00 00 32 00 00 00 68 00 74 00=·g.....·1··2··ht·
27c4b3d8	74 00 70 00 73 00 3A 00 2F 00 2F 00 77 00 77 00 2E 00 66 00 61 00 63 00 65 00 62 00 6F 00 6F 00 6B 00 2E 00 63 00 t-p-s://·w-w-w..·f-a-c-e-b-o-o-k..·c·
27c4b400	6F 00 6D 00 2F 00 6C 00 6F 00 67 00 69 00 6E 00 2E 00 70 00 68 00 70 00 20 00 5B 00 6C 00 73 00 64 00 20 00 65 00 6D 00 o-m·/·l-o-g-i-n..·p-h-p··[l-s-d··e-m·
27c4b428	61 00 69 00 6C 00 20 00 5D 00 20 00 23 00 30 00 00 00 00 10 00 00 00 00 00 00 08 00 00 00 00 00 00 00 00 0A 00 00 00 00 00 00 00 0A 00 00 00 a-i-l··]··#·0···
27c4b450	01 00 00 00 31 00 00 00 00 00 00 10 00 00 00 00 00 00 00 00 00 00 00 00 12 00 00 00 05 00 00 00 65 00 6D 00 ..1.....em·
27c4b478	61 00 69 00 6C 00 00 00 00 00 00 10 00 65 00 6D 00 a-i-l.....em·
27c4b4a0	61 00 69 00 6C 00 00 00 00 00 00 10 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 0A 00 00 00 01 00 00 00 31 00 00 00 a-i-l.....1·
27c4b4c8	00 00 00 10 00 00 00 00 00 00 00 08 00 00 00 00 00 00 00 00 00 00 00 00 16 00 00 74 00 65 00 73 00 74 00 61 00 6C 00 ..4.....t-e-s-t-a-l·
27c4b4f0	69 00 33 00 33 00 34 00 40 00 68 00 6F 00 74 00 6D 00 61 00 69 00 6C 00 2E 00 63 00 6F 00 6D 00 00 00 00 10 00 00 00 i-3-3-4@·h-o-t-m-a-i-l..·c-o-m·
27c4b518	00 00 00 08 00 00 00 00 00 00 66 00 00 00 2F 00 00 00 68 00 74 00 74 00 70 00 73 00 3A 00 2F 00 2F 00 6D 00 2E 00 ..f···f···h-t-t-p-s://·m..·
27c4b540	66 00 61 00 63 00 65 00 62 00 6F 00 6F 00 6B 00 2E 00 63 00 6F 00 6D 00 2F 00 72 00 65 00 67 00 2F 00 20 00 5B 00 6C 00 f-a-c-e-b-o-o-k..·c-o-m·/·r-e-g···[·1·
27c4b568	73 00 64 00 20 00 66 00 69 00 72 00 73 00 74 00 6E 00 61 00 6D 00 65 00 20 00 5D 00 20 00 23 00 30 00 00 10 00 00 00 s-d··f-i-r-s-t-n-a-m-e··]··#·0···0.....

Figure 2: Facebook Login Artifacts

File Content	
	Hex Text Filtered Natural
1e342708	61 72 74 65 64 2F 06 28 03 55 01 68 74 74 70 73 3A 2F 2F 77 77 77 2E 66 arted//·(·U·https://w-w-w.f
1e342720	61 63 65 62 6F 6F 6B 2E 63 6F 6D 2F 3F 73 6B 3D 77 65 6C 63 6F 6D 65 05 acebook.com/?sk=welcome·
1e342738	3F 04 81 01 01 68 74 74 70 73 3A 2F 2F 77 77 77 2E 66 61 63 65 62 6F 6F ?···https://w-w-w.facebo
1e342750	6B 2E 63 6F 6D 2F 6C 6F 67 69 6E 2E 70 68 70 3F 6C 6F 67 69 6E 5F 61 74 ook.com/login.php?login_at
1e342768	74 65 6D 70 74 3D 31 26 6C 77 76 3D 31 31 30 04 7A 04 81 77 01 68 74 74 tempt=1&lwv=110·z··w·htt
1e342780	70 73 3A 2F 2F 77 77 77 2E 67 6F 6F 67 6C 65 2E 63 6F 6D 2E 70 6B 2F 73 ps://w-w-w.goo-gle.com.pk/s
1e342798	65 61 72 63 68 3F 71 3D 66 61 63 65 62 6F 6F 6B 26 6F 71 3D 66 26 61 71 earch?q=facebook&q=f&aq
1e3427b0	73 3D 63 68 72 6F 6D 65 2E 30 2E 36 39 69 35 39 6A 36 39 69 35 37 6A 36 s=chrome.0.69i59j69i57j6
1e3427f8	74 74 70 73 3A 2F 2F 77 77 2E 66 61 63 65 62 6F 6F 6B 2E 63 6F 6D 2F ttps://w-w-w.facebook.com/
1e342810	02 6E 04 81 61 09 68 74 74 70 73 3A 2F 2F 77 77 77 2E 67 6F 6F 67 6C 65 n··a·https://w-w-w.goo
1e342828	2E 63 6F 6D 2E 70 6B 2F 73 65 61 72 63 68 3F 71 3D 66 61 63 65 62 6F 6F .com.pk/search?q=faceboo
1e342840	6B 26 6F 71 3D 66 61 26 61 71 73 3D 63 68 72 6F 6D 65 2E 31 2E 36 39 69 &og=f&aq=s=chrome.1.69i
1e342858	35 37 6A 30 6C 35 2E 32 34 30 30 6A 30 6A 37 26 73 6F 75 72 63 65 69 64 57j015.2400j0j7&sourceid

The e-mail ID "testali334@hotmail.com" is found, as shown in Figure 3. The artifacts "[https://www.facebook.com/login.php--\[l-s-d-e-m-a-i\]](https://www.facebook.com/login.php--[l-s-d-e-m-a-i])" and "[https://www.facebook.com/reg--\[l-s-d-f-i-r-s-t-n-a-m-e\]](https://www.facebook.com/reg--[l-s-d-f-i-r-s-t-n-a-m-e])" indicates that this email ID is used to register and log in to Facebook.

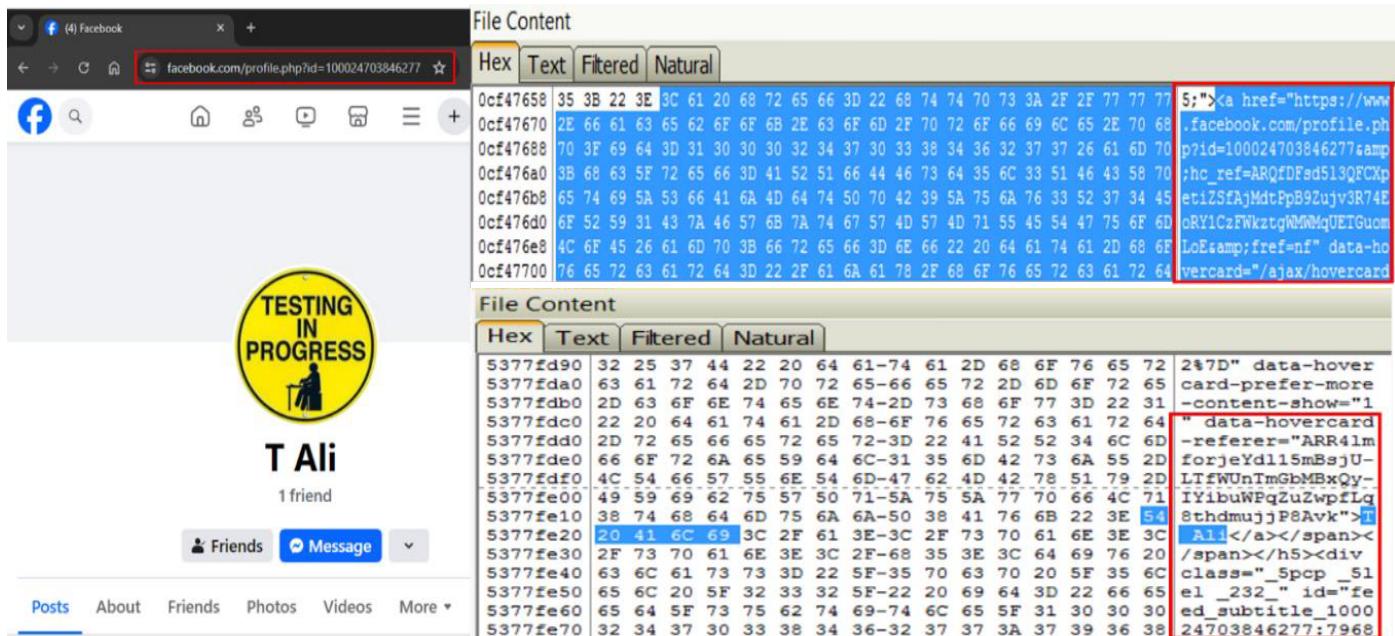


Figure 3: E-mail Artifacts of the suspect

6.2 Facebook Profile Artifacts

As email ID of the suspect is already established, it is now evident that the user's profile name also matches the email address convention. This indicates that the Facebook profile belongs to the suspected user, "T Ali". The profile URL of the alleged user "<https://www.facebook.com/profile.php?id=100024703846277>" matches the extracted URL from the forensic analysis, as shown in Figure 4. Therefore, it is established that the suspected user used this PC to log in to Facebook.

Browser forensics involves analyzing web browsers to recover digital evidence, such as browsing history, cookies, and cached files. Cookie data is crucial in this process as it can reveal user activities, login sessions, and tracking information, which help in investigations. Facebook maintains many cookies for multiple purposes such as "act", "c_user", "datr", "fr", "presence", "sb", "spin", "wd", and "xs" etc. Each cookie has a special purpose, but here we will only mention "c_user" which contains the user ID of the currently logged-in user. The cookie value "c_user=100024703846277" shown in Figure 5 is analyzed from the browser and was also found in the forensic analysis using AccessData FTK. This value refers to the currently logged-in user i.e. Facebook profile ID of the suspect, which ultimately attributes that the suspect "T Ali" was logged in to Facebook using this PC.

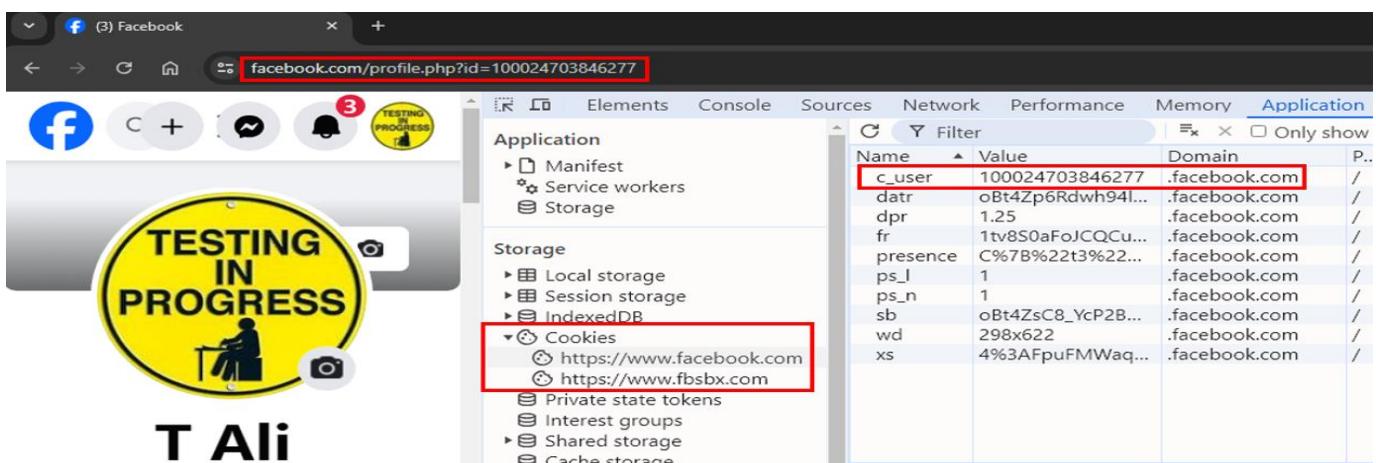


Figure 5: Cookie Data of Currently Logged-in Facebook User

6.3 Facebook Notifications Artifacts

During further analysis, the artifacts of the Facebook notification “H Ali commented on your photo” were identified. Figure 6 shows the numeric Facebook ID “100024702406643” corresponding to the Facebook user “H Ali” as shown. This confirms that “H Ali” have commented on some post.

File Content

Hex	Text	Filtered	Natural
41b98318	4D 54 55 79 4E 6A 59 34 4E 77 3D 3D 41 42 18 38 EC 5C BD 6A 14 00 00 00 31 35 32 31 34 34 30 39 33 33 30 31 38 35 32 39 A9 4D 18 38 2A F4 DC 3C 40 00 00 00 BC 94 45 56	MIUyNjY4Nw==AB·8i\>j...-1521440571...AB·8	
41b98340	0A 35 70 5B 20 00 00 00 31 35 32 31 34 34 30 39 33 33 30 31 38 35 32 39 A9 4D 18 38 2A F4 DC 3C 40 00 00 00 BC 94 45 56	-5p[...-1521440933018529@M·8*5d<@...-34·EV	
41b98366	BC AC 04 52 01 41 18 38 30 00 00 00 31 35 32 31 34 34 30 35 37 31 34 37 38 38 36 35 A9 4D 18 38 F2 33 EO 9C 78 00 00 00	\<...-R·A·80...-1521440571478865@M·8d3@·x...	
41b98390	9C 9C 45 56 64 5F 19 52 01 41 18 38 68 00 00 00 78 2E 66 62 63 64 6E 65 74 2F 72 73 72 63 2E 70 68 70 2F 76 33 2F	·EVd...R·A·8h...x.fcdn.net/rsrc.php/v3/	
41b98398	79 6A 2F 72 2F 41 4E 34 50 46 4E 52 75 6C 52 44 2E 70 6E 67 41 42 18 38 9A 21 BA CF 18 00 00 00 66 65 65 64 5F 63 6F 6D	yj/r/AN4FPNRu1RD.pngAB·8·!·I...-feed.com	
41b983e0	6D 65 6E 74 41 42 18 38 9A AA B2 32 3C 00 00 00 48 20 41 6C 69 20 63 6F 6D 65 6E 74 65 64 20 6F 6E 20 79 6F 75 72 20	mentAB·8·!*2<...-H Ali commented on your	
41b98408	70 68 6F 74 6F 2E 00 00 41 42 18 38 4E 4A 7D 66 D6 01 00 00 78 22 6E 6F 74 69 66 5F 74 79 70 65 22 3A 22 66 65 65 64 5F	photo...-AB·8NjzfO...-(notif_type:"feed_	
41b98430	63 6F 6D 65 6E 74 22 2C 22 73 75 62 74 79 70 65 22 3A 22 6F 77 6E 65 72 5F 6F 66 5F 74 61 72 67 65 74 22 2C 22 63 6F	comment","subtype":"owner of target","co	
41b98458	6E 74 65 78 74 5F 69 64 22 3A 22 31 33 32 36 37 31 38 31 34 32 33 32 39 34 36 22 2C 22 61 6C 65 72 74 5F 69 64 22 3A 22	ntext_id": "132671814232946","alert_id": "	
41b98480	31 35 32 31 34 34 30 35 37 31 34 37 38 38 36 35 22 2C 22 66 72 6F 6D 5F 75 69 64 73 22 3A 7B 22 31 30 30 30 32 34 37 30	1521440571478865","from_uids": ("100024702406643"), "microtime_s": 2406643,"content_id": "13	
41b98488	32 34 30 36 36 34 33 22 3A 22 31 30 30 30 32 34 37 30 32 34 30 36 36 34 33 27 7D 2C 22 6D 69 63 72 6F 74 69 6D 65 5F 73	ent": "1521440571478865","content_id": "13	
41b984d0	65 6E 74 22 3A 22 31 35 32 31 34 34 30 35 37 31 34 37 38 38 36 35 22 2C 22 63 6F 74 65 65 74 5F 69 64 22 3A 22 31 33	2671814232946"]...AB·8eXxx(...-Tm9oWZPchRp	
41b984f8	32 36 37 31 38 31 34 32 33 32 39 34 36 22 7D 00 41 42 18 38 A2 58 D7 78 28 01 00 00 54 6D 39 30 61 57 5A 50 63 48 52 70	b251ZXJ22XJBY3RpB25Ub2tLbjxsNTIxNDQwNlCx	
41b98520	62 32 35 54 5A 58 4A 32 5A 58 4A 42 59 33 52 70 62 32 35 55 62 32 74 6C 62 6A 73 78 4E 54 49 78 4E 44 51 77 4E 54 63 78 b251ZXJ22XJBY3RpB25Ub2tLbjxsNTIxNDQwNlCx	ISI6IkjhREU1LCJ1bmRvIjoi	

Figure 6: Facebook Notification Artifacts

6.4 Facebook Text Post and Comments Artifacts

T Ali
20 March 2018 ·

LIKE MY PICTURE?

TESTING IN PROGRESS T Ali
13 March 2018 ·

Hi, this is Test, How are you

1 comment

Like

Comment

Send

Share

H Ali

3 comments 1 share

View more comments

H Ali
pleasure

6 y Like Reply

Figure 7: Facebook Text Post and Comments

The text post “Hi, this is test, How are you” is identified in HTML format. Figure 8 illustrates the artifacts of this post but no conclusive artifacts were established to identify the author who posted.

File Content

Hex	Text	Filtered	Natural
53780520	64 69 76 3E 3C 2F 64 69 76 3E 3C 2F 64 69 76 3E 3C 64 69 76 20 63 6C 61	div></div></div><div class=" _5pbx userContent _2jv _3576" data-ft="123;">Hi, th	
53780538	73 73 3D 22 5F 35 70 62 78 20 75 73 65 72 43 6F 6E 74 65 6E 74 20 5F 32	is is Test, How are you</p></div></div><div clas	
53780550	32 6A 76 20 5F 33 35 37 36 22 20 64 61 74 61 2D 66 74 3D 22 26 23 31 32	=_5_jv _58jw"><p>Hi, th	
53780568	33 3B 26 71 75 6F 74 3B 74 6E 26 71 75 6F 74 3B 3A 26 71 75 6F 74 3B 4B	is is Test, How are you</p></div></div><div clas	
53780580	26 71 75 6F 74 3B 26 23 31 32 35 3B 22 3E 3C 64 69 76 20 63 6C 61 73 73	=_5_jv _58jw"><p>Hi, th	
53780598	3D 22 5F 35 5F 6A 76 20 5F 35 38 6A 77 22 3E 3C 70 3E 48 69 2C 20 74 68	is is Test, How are you</p></div></div><div clas	
537805b0	69 73 20 69 73 20 54 65 73 74 2C 20 48 6F 77 20 61 72 65 20 79 6F 75 3C	=_5_jv _58jw"><p>Hi, th	
537805c8	2F 70 3E 3C 2F 64 69 76 3E 3C 2F 64 69 76 20 63 6C 61 73	is is Test, How are you</p></div></div><div clas	

Figure 8: Facebook Text Post in HTML Format

JSON format can be identified on RAM. It appears to be the comment body as “body”: {text”: “I am Fine” as shown in Figure 9. Although any conclusive artifacts to establish the original author of the post and comment could not be identified.

File Content	
	Hex Text Filtered Natural
0124fc80	34 39 32 39 5D 5D 2C 5B-22 55 46 49 43 65 6E 74 4929]], ["UFICent
0124fc90	72 61 6C 55 70 64 61 74-65 73 22 2C 22 68 61 6E ralUpdates", "han
0124fca0	64 6C 65 55 70 64 61 74-65 22 2C 5B 5D 2C 5B 31 dleUpdate", [], [1
0124fcb0	32 2C 7B 22 63 6F 6D 6D-65 6E 74 73 22 3A 5B 7B 2, {"comments": [{
0124fcc0	22 62 6F 64 79 22 3A 7B-22 74 65 78 74 22 3A 22 "body": {"text": "
0124fcdo	49 20 61 6D 20 66 69 6E-65 22 2C 22 72 61 6E 67 I am fine", "rang
0124fce0	65 73 22 3A 5B 5D 2C 22-61 67 67 72 65 67 61 74 es": [], "aggregat
0124	Figure 9: Facebook Comment in “I am fine” in JSON Format , "te

In another keyword search, the artifacts for the comment “text: PLEASURE” were found. The original Facebook profile URL of the account “H Ali” is “<https://www.facebook.com/h.ali.7982780>”, further analysis using the online Facebook ID finder tool, shows the same author ID “100024702406643” as is shown in Figure 10. This indicates that the user “H Ali” commented on the post, however, the artifacts of the original post in this case were not identified.

File Content	
	Hex Text Filtered Natural
69f336d0	67 00 2F 00 31 00 33 00 32 00 36 00 37 00 31 00 38 00 31 00 34 00 32 00 33 00 32 00 39 00 34 00 36 00 2F 00 73 00 74 00 g/-1-3-2-6-7-1-8-1-4-2-3-2-9-4-6-/ -s-t-
69f336e8	61 00 72 00 74 00 2F 00 A9 43 18 38 03 00 00 00 12 00 00 00 73 6B 75 77 61 6C 68 65 72 00 00 00 A9 43 18 38 03 00 00 00 a-r-t-@C-8-----skywalker--@C-8-----
69f33720	20 0A 00 00 7B 22 63 6F 6D 6D 65 6E 74 5F 63 72 65 61 74 65 5F 73 75 62 73 63 72 69 62 65 22 3A 7B 22 63 6C 69 65 6E 74
69f33748	5F 6D 75 74 61 74 69 6F 6E 5F 69 64 22 3A 22 31 35 32 31 34 31 30 32 34 31 38 33 3A 31 36 31 36 38 32 39 38 22 2C 22
69f33770	63 6C 69 65 6E 74 5F 73 75 62 73 63 72 69 70 74 69 6F 6E 5F 69 64 22 3A 22 33 66 37 62 34 30 64 61 2D 34 62 32 30 2D 34
69f33798	88 66 64 2D 61 65 61 31 2D 36 32 36 34 62 31 30 34 32 36 62 34 22 2C 22 63 6F 6D 6D 65 6E 74 22 3A 7B 22 69 64 22 3A 22
69f337c0	69 32 39 74 62 57 56 75 64 44 6E 78 4D 7A 49 32 4E 7A 45 34 4D 54 51 79 4D 7A 49 35 4E 44 5A 66 4D 54 4D 79 4E 6A 67 31 Y29tbNWudDoxMzI2NzE4MTQyMzI5NDZfMTMyNjg1
69f337e8	4D 44 55 77 4F 44 6B 34 4D 6A 67 35 22 2C 22 6C 65 67 61 63 79 5F 66 62 69 64 22 3A 22 31 33 32 36 38 35 30 38 39 MDUwODk4Mjg5", "legacy_fbid": "13268505089
69f33810	88 32 38 39 22 2C 22 6C 65 67 61 63 79 5F 74 6F 6B 65 6E 22 3A 22 31 33 32 36 37 31 38 31 34 32 33 32 39 34 36 5F 31 33 8289", "legacy_token": "132671814232946_13
69f33838	82 36 38 35 30 35 30 38 39 38 32 38 39 22 2C 22 74 65 78 74 5F 64 65 6C 69 67 68 74 73 5F 61 72 65 5F 68 69 64 64 65 6E 2685050898289", "text_delights": are hidden
69f33860	22 3A 66 61 6C 73 65 2C 22 61 75 74 68 6E 72 22 3A 7B 22 69 64 22 3A 22 31 30 30 32 34 37 30 32 34 30 36 36 34 33 22
69f33888	:false, "author": {"id": "100024702406643", "name": "H Ali"}, "attachments": [], "body": "6F 33860", "ranges": [], "inline_style_ranges": [], "image_ranges": [], "aggright_ranges": []}, "
69f338b0	3A 7B 22 74 65 78 74 22 3A 22 50 4C 45 53 55 52 45 22 2C 22 72 61 6E 67 65 73 22 3A 5B 5D 2C 22 69 6E 60 69 6E 65 5F : "text": "PLEASURE", "ranges": [], "inline_
69f338d8	73 74 79 6C 65 5F 72 61 6E 67 65 73 22 3A 5B 5D 2C 22 69 6D 61 67 65 5F 72 61 6E 67 65 73 22 3A 5B 5D 2C 22 61 67 67 72 style_ranges": [], "image_ranges": [], "aggright_ranges": []}, "

Figure 10: Facebook Comment “pleasure” in JSON Format

7. Conclusions and Future Direction

This research highlights that forensic analysis of RAM can effectively uncover Facebook artifacts, with keyword-based searches proving reliable for reconstructing the evidence crucial in cybercrime investigations. Significant digital traces are left on RAM, offering valuable evidence for digital forensic investigators. Results can vary based on device specifications and browser combinations, and while artifacts may be found in different locations, their format remains consistent. False positives are a concern, necessitating manual verification of search results. The study used FTK Imager and AccessData FTK, but other tools like Autopsy or WinHex could also be applied. Limitations include the small size of the pagefile.sys file and evolving URL formats by Facebook, potentially impacting artifact accessibility over time. Future research should focus on developing forensic tools that accommodate localized languages and explore a range of browsers and devices. Social media forensics remains a complex and evolving field, requiring ongoing research and adaptation of open-source tools.

8. References

- [1] T. Tähtinen, "When Facebook Is the Internet: The Role of Social Media in Ethnic Conflict," *World Development*, vol. 180, p. 106633, 2024.
- [2] E. Z. Ekhato, "Attitude of UNIBEN Students Towards Facebook Usage," *Interdisciplinary Journal of African & Asian Studies (IJAAS)*, vol. 10, no. 2, 2024.
- [3] I. S. Ivasciuc, C. P. Constantin, A. N. Candrea, and A. Ispas, "Digital Landscapes: Analyzing the Impact of Facebook Communication on User Engagement with Romanian Ecotourism Destinations," *Land*, vol. 13, no. 4, p. 432, 2024.

- [4] I. Adinata and Y. Servanda, "Kajian Literatur: Metode Analisis dan Tools Live Forensics Pada Random Access Memory (RAM)," *Jurnal Sains Dan Teknologi (JSIT)*, vol. 4, no. 2, pp. 175-180, 2024.
- [5] H. Sanghvi, D. Rathod, P. Shukla, R. Shah, and Y. Zala, "Web browser forensics: Mozilla Firefox," *International Journal of Electronic Security and Digital Forensics*, vol. 16, no. 4, pp. 397-423, 2024.
- [6] K. Wong, S. Researcher, A. C. T. Lai, J. C. K. Yeung, and W. L. Lee, "Facebook Forensics Finalized," *Financial and Banking Information Infrastructure Committee (FBICC)*, [Online]. Available: https://www.fbiic.gov/public/2011/jul/Facebook_Forensics-Finalized.pdf. [Accessed: Aug. 31, 2024].
- [7] R. Thantilage and N. Jeyamohan, "A volatile memory analysis tool for retrieval of social media evidence in Windows 10 OS based workstations," in 2017 *National Information Technology Conference (NITC)*, Sept. 2017, pp. 86-88.
- [8] N. Joseph, S. Sunny, S. Dija, and K. L. Thomas, "Volatile Internet evidence extraction from Windows systems," in 2014 *IEEE International Conference on Computational Intelligence and Computing Research*, Dec. 2014, pp. 1-5.
- [9] H. Said, N. Al Mutawa, I. Al Awadhi, and M. Guimaraes, "Forensic analysis of private browsing artifacts," in 2011 *International Conference on Innovations in Information Technology*, Apr. 2011, pp. 197-202.
- [10] H. C. Chu, D. J. Deng, and J. H. Park, "Live data mining concerning social networking forensics based on a facebook session through aggregation of social data," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 7, pp. 1368–1376, 2011, doi: 10.1109/JSAC.2011.110804.
- [11] H. C. Chu, D. J. Deng, and J. H. Park, "Live data mining concerning social networking forensics based on a Facebook session through aggregation of social data," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 7, pp. 1368-1376, 2011.
- [12] J. I. James, P. Gladyshev, and Y. Zhu, "Signature based detection of user events for post-mortem forensic analysis," in *Digital Forensics and Cyber Crime: Second International ICST Conference, ICDF2C 2010*, Abu Dhabi, United Arab Emirates, October 4-6, 2010, Revised Selected Papers 2, Berlin, Germany: Springer Berlin Heidelberg, 2011, pp. 96-109.
- [13] M. S. Chang and C. Y. Chang, "Twitter social network forensics on Windows 10," *International Journal of Innovative Science, Engineering and Technology*, vol. 3, no. 9, pp. 55-60, 2016.
- [14] T. Y. Yang, A. Dehghantanha, K. K. R. Choo, and Z. Muda, "Windows instant messaging app forensics: Facebook and Skype as case studies," *PLOS One*, vol. 11, no. 3, pp. e0150300, 2016.
- [15] A. Rocha, W. J. Scheirer, C. W. Forstall, T. Cavalcante, A. Theophilo, B. Shen, ... and E. Stamatatos, "Authorship attribution for social media forensics," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 5-33, 2016.
- [16] M. Huber, M. Mulazzani, M. Leithner, S. Schrittwieser, G. Wondracek, and E. Weippl, "Social snapshots: digital forensics for online social networks," Proc. 27th Annu. Comput. Secur. Appl. Conf., pp. 113–122, 2011, doi: 10.1145/2076732.2076748.



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Article Citation:
Kanwal et al. (2024). "YOLO-Based Real-Time Recognition of Sign Language Gestures". *FUSST International Conference on Computer Science and Engineering*

YOLO-Based Real-Time Recognition of Sign Language Gestures

Tabassum Kanwal ^{a,*}, Saud Altaf ^b, Rehan Mehmood Yousaf ^a

^a University Institute of Information Technology, PMAS-University of Arid Agriculture, Rawalpindi, 46000. (tabassum.kanwal@iiu.edu.pk). (rehan.yousaf@uaar.edu.pk)

^b Department of Information Engineering Technology, National Skills University, Islamabad 44000, Pakistan; (saud.altaf@nsu.edu.pk).

* Corresponding author: tabassum.kanwal@iiu.edu.pk

Abstract:

The most fundamental form of communication for the deaf and hard of hearing is sign language (SL). But most healthy persons are unaware of these SLs. Our research introduces a sign language translator that can accept a sign gesture as input and output it on a display. For the system's training, we utilized Convolutional Neural Networks (CNNs) using a provided database. The Pakistani SL database, which has 26 letters and 10 numbers, was utilized in our work. For picture segmentation, we used the Histogram Back Projection method. A convolutional neural network (CNN) is trained and tested using these datasets once they have been organized into classes. Finding a testing accuracy of 99.89% and a validation accuracy of 99.85% after 50 epochs after training feels great. The next step is to put the system through its paces using real-time data and see the outcome on the screen.

Keywords: Sign Language Recognition, Real-Time Sign Language Translation, Convolutional Neural Networks (CNNs), Gesture Recognition.

1. Introduction

In order for computers to show human gestures or take them as orders, gesture recognition—a form of non-cognitive computing user interface—captures and interprets the movements. There is a lot of activity in this area of computer vision research, mostly related to HCI. Those who are hard of hearing and rely on hand gestures as their only means of communication have found this technology to be of tremendous assistance. The World Health Organization estimates that 63 million individuals in India have hard of hearing (2018) [1]. This community uses Pakistani Sign Language (PSL) as their primary means of communication. India and Bangladesh are only two of the numerous South Asian nations where PSL is the de facto sign language. made up of dynamic and static gestures as shown in figure 1. About three thousand words make up PSL. The most basic form of communication for those who are hard of hearing or deaf is sign language (SL).

A major obstacle to communication arises, however, because the majority of people are unfamiliar with these gestures. One promising avenue for addressing these issues and facilitating communication between the hearing and the deaf is the creation of real-time sign language recognition systems.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

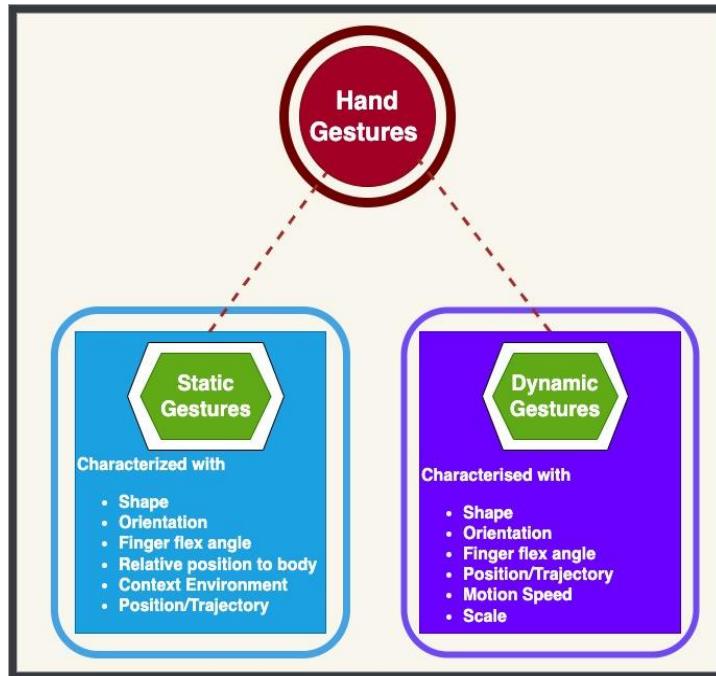


Figure 1: Types of gestures

The social advantages of this system would be enormous, considering that 63 million individuals in sub-Continent were estimated to have hearing loss in 2018 according to the World Health Organization. This project investigates the use of deep learning techniques, specifically Convolutional Neural Networks (CNNs), to develop a sign language gesture translator that can dynamically convert motions into text or speech in real-time. In addition to enhancing accessibility, this approach promotes diversity in multiple sectors, including customer service, education, and healthcare. A more inclusive society with fewer communication barriers is the goal of the proposed technology, which uses Pakistani Sign Language (PSL) to provide near-instantaneous translation.

The following is the paper's structure: Section II reviews relevant literature, Section III lays out our plans, Section IV reviews the outcomes of our experiments, and Section V ends our work.

2. Related Work

There has been a lot of research into systems that can recognize hand gestures and sign languages. However, there is limited work done based on deep learning. An article on ISL gesture recognition that makes use of deep learning and image processing was published by Bhagat et al. [2]. They, too, have trained 36 static motions corresponding to the letters and digits in ISL using CNNs. They were able to accomplish a one-to-one mapping between the depth and the RGB pixels by utilizing computer vision algorithms. Five convolutional layers made up their CNN model. On training photos, the suggested model reached 98.81% accuracy. Training 10 ISL dynamic word gestures using Convolutional LSTMs yielded a 99.08% success rate on 1080p films. Intwala et al. [3] also published a work that we consulted; in it, the authors use MATLAB to generate their own dataset, which they then pre-process using several methods. They then sent the information into the AlexNet classifier. The GrabCut method is employed for the purpose of segmentation. The model is tested for each unique letter and the average test accuracy for the real-time photos is determined to be 87.69%. In one of their published papers, Sahoo et al. [4] demonstrated hand gesture identification with the use of an SVM classifier and Deep Convolutional Neural Networks (DCNs) that were reduced using Principal Component Analysis (PCA). They have tested hand motions using AlexNet for feature extraction. American Sign Language is the primary focus of their work, and they have utilized many test participants to construct databases for this purpose. With their suggested approach, they were able to achieve a test accuracy of around 99.32%. Sarkar et al. [5] also used CNNs to create a comparable system.

Additionally, they have built their very own database. It all starts with a camera feeding the input image into the CNN model. Additionally, this system uses data that is updated in real-time to generate output. A camera, though, is an external equipment that must be used for this project. This model has an accuracy of around 99.40%.

Sruthi and Lijiya [6] discuss this in their study. A technique for ISL recognition based on deep learning that is independent of signers is suggested in their work. They built the database using a vision-based recognition system that processes images using two algorithms: one for skin color segmentation and another for hand region segmentation. A CNN model is subsequently trained on the data. They achieve a training accuracy of 99.93% and a testing and validation accuracy of 98.64% utilizing 24 ISL static alphabets.

We built our dataset from an image processing model that uses histogram backpropagation for hand segmentation and detection as part of our study. Afterwards, a convolutional neural network (CNN) model equipped with a sequential classifier is given this data. In response to input in real time, this system displays an output in a separate window.

3. Proposed Method

In our proposed system, there are different phases in the framework. The main methodology of our model, which shown in Figure 2.

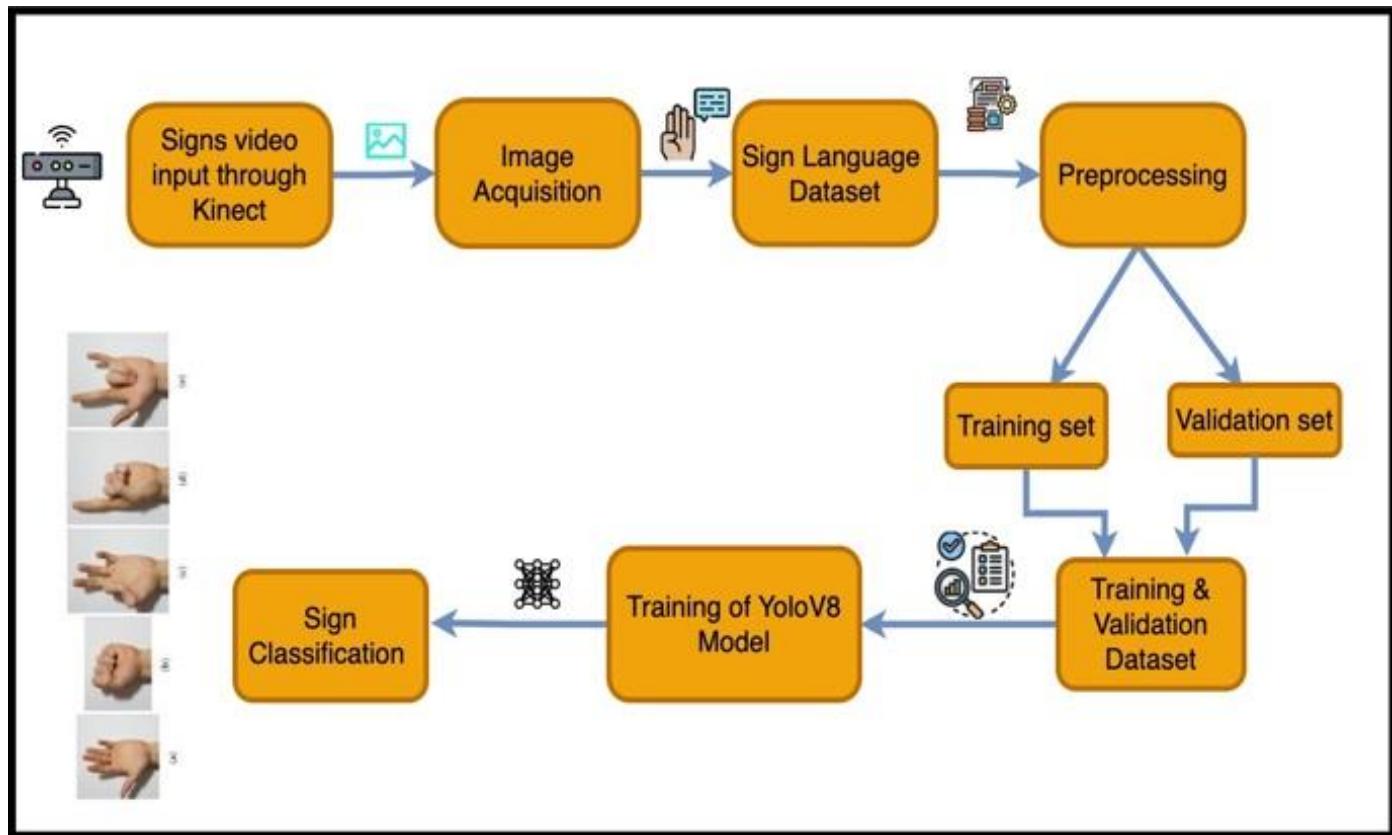


Figure 2: Block diagram of the proposed system

Each block is discussed one by one below:

3.1. Hand Segmentation and Detection

To provide precise gesture identification in real-time applications, hand segmentation is an essential system step. The system's accuracy in detecting and isolating the hand from the backdrop can be greatly affected by different lighting conditions, which is one of the main obstacles in this procedure. Traditional segmentation methods struggle to isolate the hand in images with bright or uneven lighting because of the resulting shadows, reflections, or color fluctuations. Movement artifacts, backdrop clutter, and skin tone differences all add more difficulty to the endeavor. We overcame these obstacles by employing the Histogram Back Projection technique, which works wonders in situations with unpredictable and dynamic illumination for locating the hand. In order for this technique to function, a color histogram of the hand is generated using the HSV color space. This space is more resistant to changes in lighting than the RGB color system. Transforming the hand's color data from RGB to HSV allows the system to zero in on saturation and hue, mitigating the effects of lighting-induced brightness variations. Even in low-light situations, the histogram can improve segmentation accuracy by drawing attention to the image pixels most likely to represent the hand.

We apply morphological procedures and noise-removal filters to further improve the outcome after manual segmentation using back-projection of the histogram. Whether or whether the user is wearing gloves, these post-processing procedures smooth out edges and remove minor artifacts from the segmented image, resulting in a clean hand silhouette. In our tests, we found that the system consistently performed well under a wide range of illumination circumstances, proving that this method is quite versatile and robust.

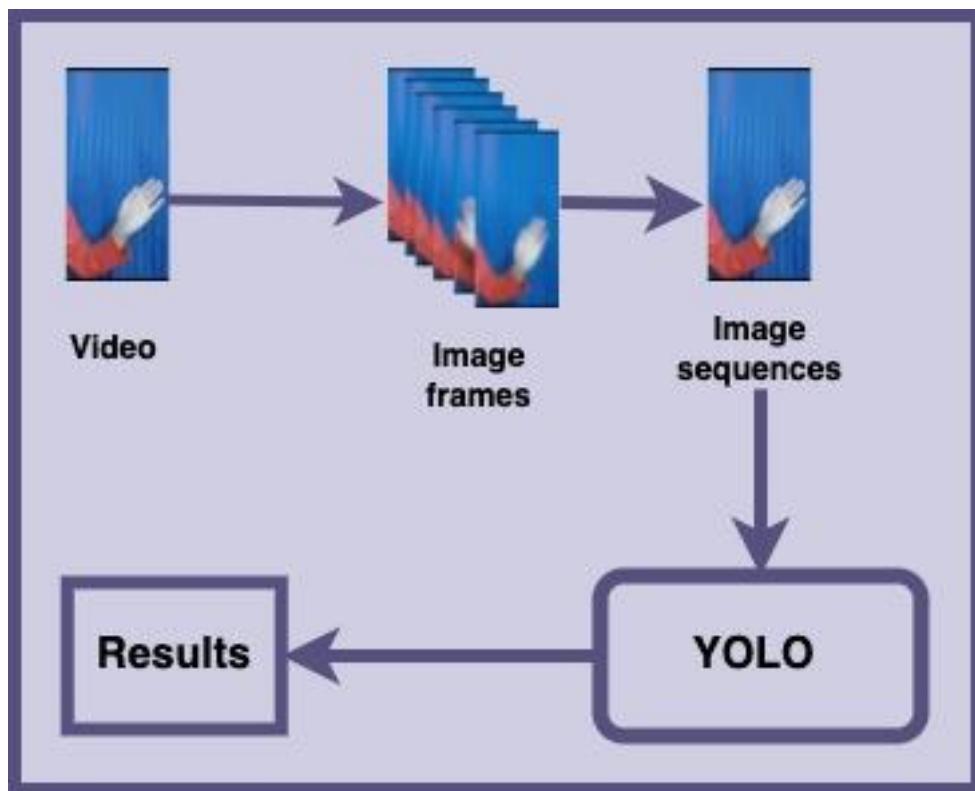


Figure 3: Flow for hand detection

Figure 3 shows the processes to acquire the image's histogram. This proves that the segmentation is effective against bright backgrounds whether the user is wearing gloves or not. We have continued the experiment while wearing gloves as that is how we obtain the best results.

3.2. Database

A vast dataset is required to get an accurate model. Since no active dataset is currently accessible, we discovered that they had developed their own dataset by comparing it to previous research. Because of this, we have also built our own database. Using the aforementioned method, we were able to obtain 2400 photos for every motion after flipping 1200 of them. We have gathered ten numbers and twenty-six letters from the PSL script.

We need to make a window on the screen that the user may place their hand in before we can record their action. Once again, we have shown the threshold in a separate window, and we load the histogram into that area to make it.

sorting and morphological processing is carried out. After that, we attached a timer to the video feed so that it may record periodically. After we launch the software, we set the timer to take up to 1200 photos sequentially. In Figure 4, we can see one of the recorded frames. The picture will be 50 pixels wide and 50 pixels tall. While the frames are being taken, we can also adjust the hand gesture's orientation. Once all the motions have been recorded, the pictures are inverted. In all, there are 72,000 photos of a 50x50 pixel size. Figure 5 shows the further partitioning of this dataset into testing, training, and validation sets. Our system's ability to include any sign language or gesture is one of its many benefits. Simply inputting the name or text of the motion will do the trick.

3.3. CNN base YOLO Architecture

A CNN base YOLO will be used for both the training and testing processes. A sequential classifier[7] was employed in the development of our model. In Figure 4 below, you can see the CNN base YOLO architecture that we have developed.

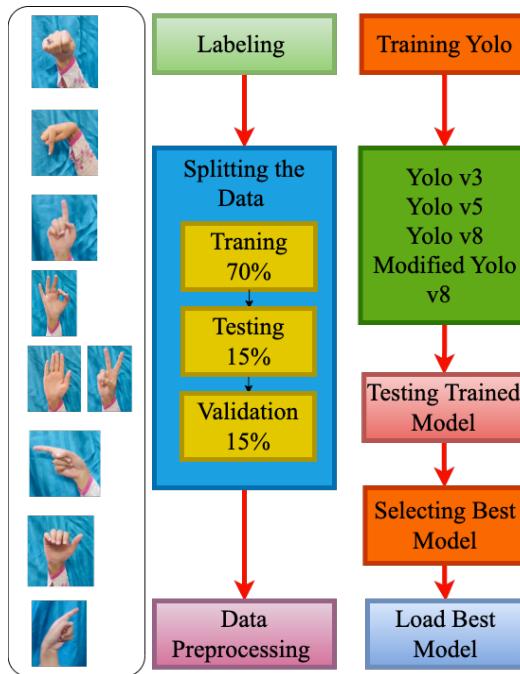


Figure 4: Architecture of the CNN base YOLO model

The initial convolutional layer receives the input image. A Rectified linear unit (ReLU) is then used to process this. THE ReLU is an activation function that preserves positive values and maps negative ones to zero, allowing for more efficient and quicker training. A Max Pooling layer performs nonlinear down sampling on the output. This aids in decreasing the number of parameters that the network has to learn. The array is flattened into a single dimension in the flatten layer, which is part of the fully connected (FC) layer that follows. The second layer is the dense layer, which is characterized by highly interconnected neurons. Stochastic gradient descent (SGD) is the optimizer we employed in this case [8].

4. Experimental Results

As seen in Figure 5, after the dataset is complete, we can verify it by displaying a single gesture for each letter A–Z and the numbers 0–9. The dimensions of these pictures are 50X50.



Figure 5: A display of the stored data

Then, we try training the model at various epochs to see if we can get the best possible outcome. We executed the program in four distinct epochs to achieve this goal. Our findings from the table indicate that, up to a certain point, the accuracy improves as the number of epochs grows. Overtraining the model causes it to drop thereafter. We get the best possible outcome in 50 epochs. As the number of epochs rises, so does the computational time. Figure 6 shows the training and validation accuracy plot over 50 iterations.

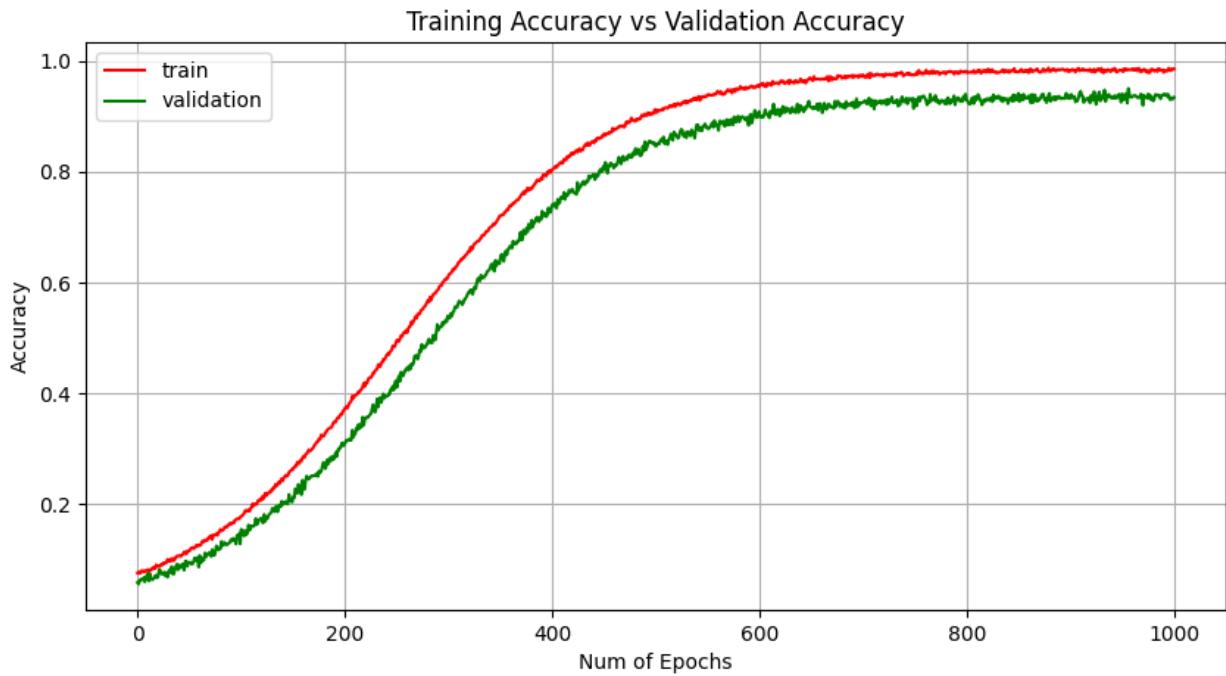


Figure 6: Training and validation accuracy with epoch 50

A. Training with live input

We put our hand in the box and hit the key to activate the CNN model to see how the system performed. what happens when it activates: a new window appears, and it displays the desired indication.

B. Comparison with other methods

From the four reference papers, that we have studied, a comparison is made on their techniques with their dataset and accuracy with our model is shown in figure 7.

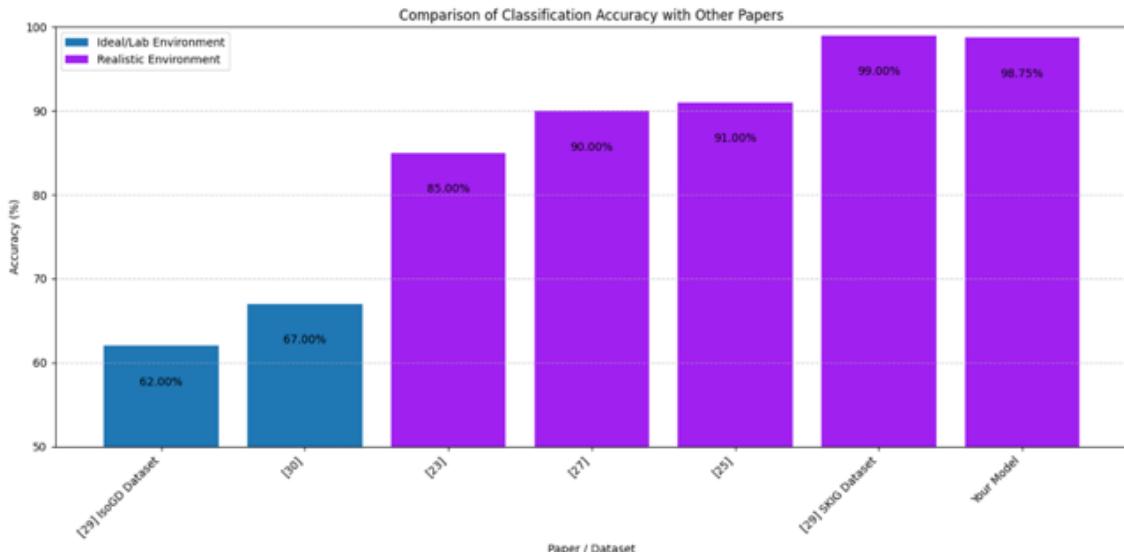


Figure 7: Comparison with other models

5. Limitation

Despite the impressive accuracy of our system in sign language identification, it is crucial to consider the dataset's constraints in order to assess how well the system performs in real-world scenarios. A large portion of the Pakistani Sign Language (PSL) standard hand gesture collection is utilized in this project. Unfortunately, the wide range of user-specific hand shapes and sizes was overlooked. Less accurate recognition findings may be experienced by individuals with larger, smaller, or uniquely shaped hands, for example. Given that the system was trained on a very uniform set of hand forms, it may have difficulty generalizing when faced with hands that are drastically different from the training data. The system's ability to handle diverse user demographics could be enhanced by adding a broader range of hand shapes to the dataset.

The lack of variation in skin tone is another shortcoming of the dataset. Varieties in skin tone may impact the model's accuracy as the hand segmentation approach is primarily focused on color-based approaches, such as Histogram Back Projection in the HSV color space. Different levels of segmentation accuracy may be observed for skin tones that are lighter or darker, particularly in difficult lighting situations. Extreme variations in skin tone, particularly in low-light conditions, may compromise the accuracy of hand detection, even though the HSV color space is more light-resistant than the RGB space.

In order to make segmentation easier, the training dataset mostly included photographs taken against solid backgrounds. However, users may be required to execute motions in intricate settings with distracting backdrops in real-world scenarios. The accuracy of the system in non-laboratory conditions may be compromised due to the limited incorporation of such variances in the training data. Future datasets should include motions done in a wider variety of contexts, such as outdoors and cluttered indoors, to increase the system's resilience.

6. Conclusions

In this research, we introduced a method for recognizing sign languages using static hand gesture movements. We have utilized ISL for the purpose of our case study. This model has set the image's histogram using the Back Projection Histogram technique. We achieved a test accuracy of 99.89% by training and testing with CNNs. Our model's independence from any third-party hardware or gadget is one of its main selling points. Background light with good lighting conditions is the only need. Furthermore, it is just useful for motionless motions. Future work can involve stringing these letters together to make meaningful words, further simplifying communication.

Acknowledgment

We are thankful to the research reviewers who helped in writing and improving these guidelines.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

7. References

- [1]. Elakkiya R (2021) Machine learning based sign language recognition: a review and its research frontier. *J Ambient Intell Humaniz Comput* 12:7205–7224
- [2]. Shukor AZ, Miskon MF, Jamaluddin MH, Bin Ali F, Asyraf MF, Bin Bahar MB (2015) A new data glove approach for Malaysian sign language detection. *Proc Comput Sci* 76:60–67
- [3]. Ren Z, Yuan J, Meng J, Zhang Z (2013) Robust part-based hand gesture recognition using Kinect. *IEEE Trans Multimed* 15(5):1110–1120
- [4]. Naglot D, Kulkarni M (2016) Real time sign language recognition using the leap motion controller. In: International conference on inventive computation technologies (ICICT)
- [5]. Almeida SGM, Guimaraes FG, Ramírez JA (2014) Feature extraction in Brazilian sign language recognition based on phonological structure and using RGB-D sensors. *Expert Syst Appl Int J* 14(6):7259–7271
- [6]. Khomami SA, Shamekhi S (2021) Persian sign language recognition using IMU and surface EMG sensors. *Measurement* 168:108471
- [7]. Tateno S, Liu H, Ou J (2020) Development of sign language motion recognition system for hearing-impaired people using electromyography signal. *Sensors* 20(20):5807
- [8]. Rastgoo R, Kiania K, Escalerab S (2021) Sign language recognition: a deep survey. *Expert Syst Appl* 164:113794

Article Citation:
 Ain et al. (2024). "Simulation-based Analysis of Materials for Artificial Axon using COMSOL". *FUSST International Conference on Computer Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Simulation-based Analysis of Materials for Artificial Axon using COMSOL

Qurat Ul Ain ^a Dr Ateeq Ur Rehman ^{b *}

^a Department of Biomedical Engineering Technology, Foundation University, Islamabad, Rawalpindi, Pakistan.
 (quratulainkiani111@gmail.com).

^b Department of Biomedical Engineering, Koc university, Istanbul, Turkey
 (arehman15@ku.edu.tr).

* Corresponding author: arehman15@ku.edu.tr

Abstract:

A COMSOL-based comparative analysis was conducted to evaluate the suitability of *gold, platinum, iridium, and aerogel* as potential coating or membrane materials for artificial axons, considering their functional performance and biocompatibility. A comprehensive literature-based analysis is carried out for polyaniline. Finally, a comparison has been made between simulation-based and literature-based analysis of materials for axons.

Keywords: Artificial Nerve; Tissue engineering; COMSOL Multiphysics simulations; biomaterials; gold; Polyaniline; platinum; iridium; aerogel;

1. Introduction

Neurons are excitable cells responsible for action potential in the nervous system causing the electrical signal to be transmitted. Neurons communicate via synapses which are gaps between consecutive dendrites and axons. [1] Neurons, the fundamental units of the nervous system, lack the ability to regenerate themselves after being injured or destroyed. When neurons are lost, they result in permanent functional deficits as seen in various neurodegenerative diseases. Research has made strides in understanding the complexity of neurons, but it still remains a question mark when it comes to the treatment of these deadly neurodegenerative diseases. Diseases such as Amyloidoses (Jacob's disease, Alzheimer's disease, dementia), Tauopathies (chronic traumatic encephalopathy, Pick's disease, progressive supranuclear palsy), Synucleinopathies (Lewy body disorders, multiple system atrophy), TDP-43 Proteinopathie (amyotrophic lateral sclerosis, progressive muscular atrophy), and many more, cause permanent neural dysfunction. [2]

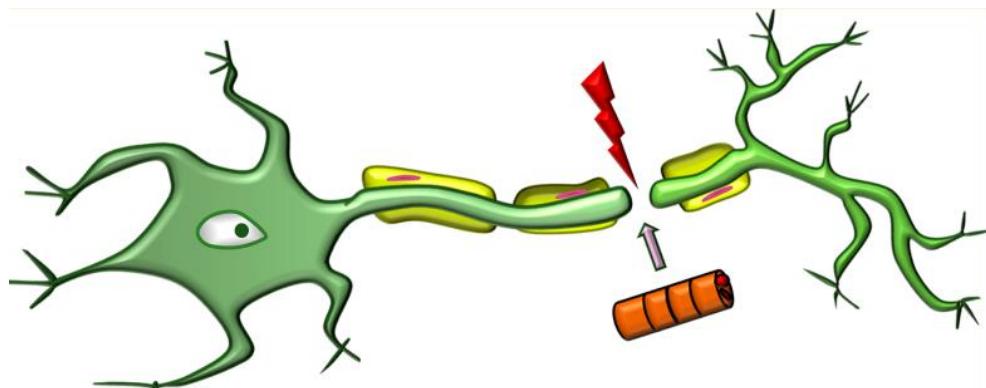


Figure 1: Peripheral nerve injury [3]

Artificial nerve synthesis has become a hot topic due to the incapability of neurons to regenerate themselves. Scientists have been working on neuron regenerations via conduits, grafts, and gene therapies, surgical techniques, electrical stimulations but still there is no single universal promising result that could be implemented for neuronal regeneration. Studies have been made on different materials such as spider silk fibers [4], fibrin matrices [5], polylactic acid [3] for the synthesis of artificial nerve conduits but their

biocompatibility and strength is still a considerable aspect and requires further research. Analyzing dozens of materials in labs can be a challenging task and requires proper apparatus and conditions. In order to save time and to minimize unnecessary animal testing, we propose a simulation-based analysis of biomaterials for artificial axon synthesis using COMSOL Multiphysics.

1.1. Research Objectives

The development of artificial axons using biomaterials has the potential to revolutionize neural prosthetics, brain-machine interfaces, and treatment of neurodegenerative disorders. Simulation-based analysis will be used to analyze



Published by
 Foundation University
 Islamabad.
 Web: <https://fui.edu.pk/>

and predict the behavior of these complex systems, accelerating the development of artificial neurons and creating more effective and reliable neural interfaces.

1.1.1. Research Objective 1

To analyze the physical, electrical, and mechanical properties of candidate materials for artificial nerve synthesis.

1.1.2. Research Objective 2

To investigate the scalability and manufacturability of biomaterial-based artificial axons using simulation-based approaches.

1.1.3. Research Objective 3

To use literature-based material evaluation approach

2. Related Work

Researchers have been developing artificial nerve guides for peripheral nerve regeneration, exploring biocompatible and biodegradable materials for a two-layer guide. They investigated various materials, including poly(caprolactone) (PCL), chitosan (CS), and poly(ester-urethane) (PU), as well as natural polymers like gelatin (G) and poly(L-lysine) (PL) for internal coatings. The study aimed to determine the optimal material combination for nerve guides, evaluating factors like biocompatibility, degradation rate, and nerve regeneration quality through in vitro and in vivo tests. [6]

A 2008 German study investigated using spider silk fibers to aid in nerve regeneration in rats with sciatic nerve injuries. The study found that nerve regeneration was successful using spider silk, either alone or with added cells or gel, but not in control groups without spider silk. The spider silk promoted Schwann cell migration, proliferation, and axonal regrowth, leading to well-aligned and healthy axons. This suggests spider silk is a promising material for peripheral nerve regeneration. [7]

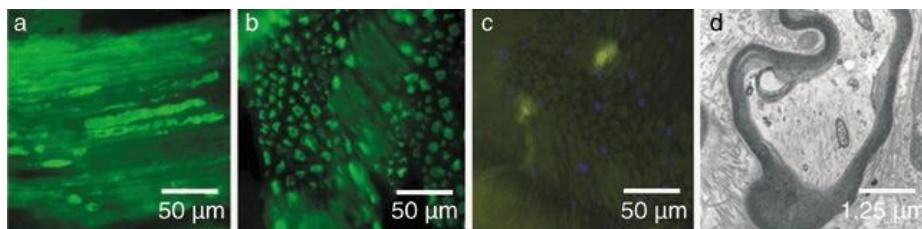


Figure 2: Axonal regeneration in a conduit consisting of vein, spider silk, Schwann cells and matrigel. . [7]

A study on sheep investigated the use of spider silk fibers for peripheral nerve regeneration, comparing them to autologous nerve grafts. After 10 months, the spider silk constructs showed successful axonal regeneration, myelination, and Schwann cell migration, with no significant difference in electrophysiological results compared to the control group. The study demonstrates the effectiveness of spider silk in enhancing nerve regeneration, leading to functional recovery. These findings have significant clinical implications for reconstructive nerve surgery, offering a viable alternative to traditional nerve transplantation methods [8].

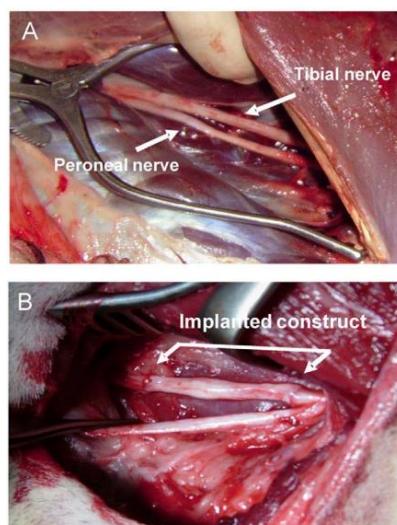


Figure 3: Imaging of peripheral nerve lesion and construct implantation in adult sheep: (A) tibial and peroneal nerve identification, and (B) 6 cm tibial nerve defect bridged with vein/spider silk construct. [8]

Researchers have created optoelectronic artificial efferent nerves using light-emitting memristors (LEMs), which combine light reception, emission, and synaptic functions to overcome limitations in electronic neural devices. This

innovation enables advanced features like one-to-many transmission and dynamic adjustable transmission, paving the way for advancements in sensorimotor functionalities and offering a promising solution for developing artificial optoelectronic nerves. [9]

A recent study examined the effect of different artificial nerve graft (ANG) structures on repairing rat sciatic nerve defects, offering insights for future ANG design. The study found that ANGs with internal microchannels promoted nerve regeneration, while those with micron-sized pores hindered the process. The findings suggest that careful design of ANG structures is crucial for effective peripheral nerve repair and highlight the importance of internal microchannels in enhancing the regeneration process. [3]

Artificial nerve guides (NGCs) are being developed to enhance peripheral nerve regeneration, with researchers exploring various materials including synthetic polymers, natural materials, and hydrogels. To overcome limitations such as mechanical instability and lack of growth-promoting signals, new materials and designs like biodegradable polymers, piezoelectric materials, and hydrogel-based NGCs are being developed, along with enhanced acellular tissues. [10]

Another study successfully fabricated artificial nerves using tissue engineering methods. Schwann cells were cultured and seeded on polyglactin 910 scaffolds coated with biomembrane, rat-tail glue, and laminin. Microscopic observations showed that Schwann cells migrated, proliferated, and formed a Bungner band on the scaffolds, producing matrices. In an *in vivo* rabbit model, the tissue-engineered conduit guided axonal regeneration in the distal nerve stump at 8 weeks. The study demonstrated the feasibility of producing adult Schwann cells on coated fibers and biomembranes, creating three-dimensional scaffolds with the basic characteristics of artificial nerves. This approach offers a promising method for fabricating tissue-engineered artificial nerves to repair long nerve defects. [11]

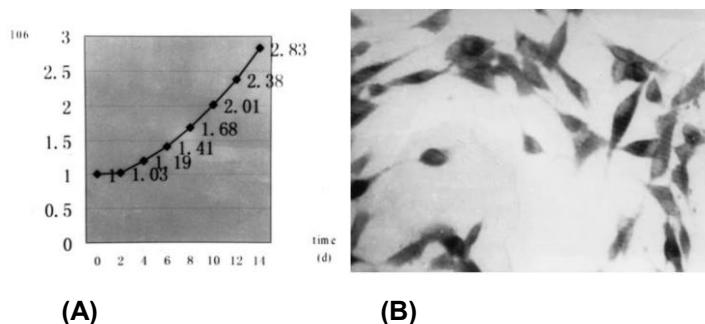


Figure 4: A) Some areas having bungner bands while others having clusters B) The SC attached to the fiber has a round nucleus, abundant cell plasma, and well-developed organelles. [11]

3. The Research Problem

Despite sustained research efforts, a single solution for neurodegenerative diseases like Guillain Barre Syndrome, Alzheimer's disease, and epilepsy remains elusive. In Pakistan, artificial nerve conduits are not available and prohibitively expensive when imported, leaving patients to face mortality or paralysis despite extensive treatment. Furthermore, the reliance on animal testing for nerve regeneration research raises ethical concerns, as most experiments are unsuccessful and result in animal fatalities. Therefore, simulation-based software studies should be prioritized, aligning with recommendations from WHO, FDA, and other regulatory bodies to leverage literature and computer simulations for biocompatibility assessments.

4. Proposed Solution

We propose a simulation-based analysis of materials for artificial axons, neurons or complete nerves using COMSOL Multiphysics software. We have used potentially suitable materials with proven biocompatibility to assess their mechanical and physical properties. Materials are assessed for their response to forces from different axes, their electrical conductivity, and temperature response. In addition, a literature-based evaluation was also performed to assess biocompatibility of chosen materials. A total of 5 materials were analyzed: 4 with COMSOL and 1 with literature-based strategy. Literature-based validation of materials was performed using renowned databases such as PubMed, Google Scholar, and FDA database. Finally, we have compared the two strategies and results to choose the most suitable material for the synthesis of artificial axons which will not only conduct electrical signal but can also withstand stress and pressure.

5. Experiments

We have used two ways to analyze these materials for artificial axon synthesis. The first 4 materials mentioned above were analyzed using COMSOL Multiphysics while the last one was analyzed only through literature.

Table 1: Experiment outline

Aspect	Details
Geometry	A 3D model of an axon was created in COMSOL, using a 500 μm long cylinder with a 4 μm radius to mimic actual axon dimensions.
Materials	Simulations were conducted using Gold, Platinum, Iridium, and Aerogel for the artificial axon synthesis.
Physics	Three COMSOL modules—Electrical Currents, Bioheat Transfer, and Solid Mechanics—were used for accurate material analysis.
Results	The analysis focused on electric potential, stress, and temperature response after building a fine-grain mesh of the axon.

1. Gold

Gold is one of the most biocompatible metals. [12] Evolving medicinal uses of gold and its compounds have spanned centuries, from ancient shamanic practices to modern pharmaceuticals. Gold salts and radioisotopes are pharmacologically valuable, with anti-inflammatory properties utilized in treating arthritis and other conditions. Additionally, gold alloys are employed in dentistry, while colloidal gold and gold isotopes have applications in research, immunogold labeling, and nuclear medicine for diagnostic and therapeutic purposes. [13]

2. Platinum

In research settings, platinum filaments serve as conductive elements; platinum crucibles and fixtures are employed in thermal decomposition studies due to their uncompromising resistance to chemical reactivity when exposed to extreme thermal conditions (approaching 1000 °C). Platinum is utilized as a modifying agent to enhance the properties of various metallic products, medical tools, dental implants, electrical connectors, and temperature sensors. Platinum-based pharmaceuticals are used to combat a diverse range of malignancies, including testicular and ovarian tumors, skin cancer, lung cancer, bone marrow cancer, and immune system cancer. [14]

3. Iridium

Iridium has been utilized as a material for neural interfaces due to its exceptional properties, making it an ideal choice for neural microelectrodes. The use of iridium oxide (IrO_x) in conjunction with platinum (Pt) has been shown to significantly enhance the performance of neural microelectrodes. [15] Furthermore, iridium exhibits excellent biocompatibility and chronic stability, making it suitable for long-term implantation in the central nervous system. The modified microelectrodes have been shown to enhance the adhesion of microglia, the primary immune cells in the brain, indicating a high level of biocompatibility.

Iridium's biocompatibility is due to its:

- Low toxicity
- Resistance to corrosion and oxidation
- Inertness to body fluids and tissues
- Ability to withstand high temperatures without releasing toxic substances

4. Aerogel

Aerogel is being explored as a promising drug delivery system [16] due to its exceptional biocompatibility and unique structural properties. Its high surface area and porous nature enable efficient adsorption of drugs from supercritical CO₂, allowing for controlled release. The release rate of drugs can be tailored by modifying the aerogel's properties, such as pore size, surface chemistry, and density. This enables:

- Sustained release
- Targeted delivery
- Improved bioavailability

Aerogel's biocompatibility ensures minimal tissue irritation and toxicity, making it an attractive platform for drug delivery applications. The use of supercritical CO₂ for drug loading enhances efficiency and scalability. Overall, aerogel shows potential as a versatile and effective drug delivery system for various therapeutic applications.

We analyzed aerogel for its potential use in artificial nerves because we were drawn to its exceptional properties. We believed these properties could be harnessed to create a groundbreaking approach to nerve regeneration and repair.

As we explored the field of biomaterials and tissue engineering, we became increasingly convinced that aerogel's high porosity, surface area, and biocompatibility makes it an ideal candidate for supporting nerve cell growth and function.

We were particularly interested in using aerogel as a scaffold for nerve tissue engineering, as it seemed to offer a promising solution for bridging the gap between damaged nerve ends and restoring motor function.

Through our analysis, we aimed to explore the potential of aerogel to:

- Support nerve cell attachment, proliferation, and differentiation
- Facilitate nerve signal transmission and conduction
- Enhance nerve regeneration and repair
- Develop a biocompatible and conductive material for neural implants and prosthetic interfaces

By investigating aerogel's properties and potential applications in nerve tissue engineering, we sought to contribute to the development of innovative solutions for treating nerve damage and neurological disorders, ultimately improving the lives of individuals affected by these conditions.

Detailed results and pictures are available in the appendix.

6. Literature-based analysis of Polyaniline

Polyaniline (PANI)

Conductive polymers, such as polyaniline (PANI), offer unique electro-optical properties and significant potential in biomedical applications due to their exceptional electrical conductivity, biocompatibility, and stable nanostructure, with their limitations overcome through strategic blending and nanocomposite formation. A review paper provides a comprehensive examination of the cutting-edge biological activities and applications of PANI-based conductive nanocomposites in biomedical domains, including antimicrobial therapy, drug delivery systems, biosensors, nerve regeneration, and tissue engineering. By scrutinizing recent advancements in PANI-based nanocomposites, this review aims to furnish a foundation for future research endeavors. [17]

Studies have shown that PANI exhibits biocompatibility properties, including dermal irritation and sensitization, according to ISO 10993 standards. However, cytotoxicity can arise from low-molecular-weight compounds. Modification of PANI is essential to mitigate these effects. Research has demonstrated that PANI-based materials can enhance cell proliferation, showing promise for tissue regeneration. [18] Investigations into the biocompatibility of PANI cryogel [19], colloidal PANI, and PANI nanofibers have yielded varying results. Factors such as chemical composition, size, shape, dopants, and preparation methods influence biocompatibility. PANI biocomposites, like starch/PANI, have shown increased biocompatibility with higher naturally occurring polymer content. [20]

Polyaniline (PANI) is a promising material for nerve regeneration and cardiac tissue engineering due to its electrical conductivity, biocompatibility, and environmental stability. Recent studies have explored the incorporation of PANI into bioactive scaffolds for tissue engineering applications, demonstrating its potential to electrically stimulate cells and influence tissue regeneration. Conductive PANI/PEGDA [21] macroporous hydrogels have been developed for nerve regeneration, showing enhanced morphological and conductive properties. PANI-based scaffolds [22] have also shown potential in supporting neurite outgrowth and promoting neuronal adhesion and extension. In cardiac tissue engineering [23], PANI and its blends have been investigated for their electrical conductivity and biocompatibility, with PANI-based cardiac patches showing high biocompatibility and conductivity. Additionally, electrospun [24] silk-polyaniline conduits have been used for functional nerve regeneration in a rat sciatic nerve injury model, demonstrating safety and efficacy. However, the biocompatibility and biodegradability of PANI need to be further investigated for long-term implantation success. Overall, PANI-based materials show great potential for biomedical applications, particularly in nerve regeneration and cardiac tissue engineering. [25] [26] [27] [28]

7. Results and Discussion

A simulation-based analysis of four materials; Gold, platinum, Iridium, and Aerogel, revealed their strength, temperature response, and electrical conductivity. Gold, renowned for its biocompatibility, exhibited promising results in all three aspects, making it a potential candidate for coating artificial nerve conduits. Platinum and iridium showed favorable temperature responses, but their electrical conductivity and strength varied. Platinum demonstrated good electrical conductivity and strength, while iridium's strength was lower compared to gold and platinum. Aerogel, a non-conductive substance, can be used as an insulating material.

Table 2: Stress analysis

Axes			Max Nm ⁻²				Min Nm ⁻²			
x	y	z	Gold	Platinum	Iridium	Aerogel	Gold	Platinum	Iridium	Aerogel
10	0	0	700	700	7	60	100	100	1	10
0	10	0	2×10^4	3.5×10^4	6	70	0.5×10^4	0.5×10^4	1	10
0	0	10	2×10^4	3.5×10^4	6	80	0.2×10^4	0.5×10^4	1	10

The simulation results underscore the significance of material selection in biomedical applications, particularly for gold, platinum, iridium, and aerogel. Notably, gold's high ductility and conductivity make it an ideal choice for implantable devices, such as pacemakers and stents, whereas platinum's high strength-to-weight ratio suits it for structural components in biosensors. Iridium's exceptional corrosion resistance and biocompatibility render it suitable for long-term implantation, while aerogel's unique porous structure shows promise for tissue engineering scaffolds. However, this research solely focuses on their potential for nerve axon regeneration or conduits, exploring how these materials' properties can enhance neuronal growth and guidance. Future studies can investigate their broader applications in biomedical engineering. Simulation-based studies enable the visualization of material properties, facilitating informed selection and testing. This approach supplements literature-based evaluations, which rely on previous studies to determine material properties, biocompatibility, and functionality. For novel materials, literature reviews should precede animal testing to ensure comprehensive understanding. Unnecessary animal testing should be avoided, as it poses risks to animals and contributes to biodiversity loss.

8. Conclusions and Future Direction

Our research represents a pioneering step towards the development of artificial neurons or complete nerves, paving the way for groundbreaking advancements in the field. Through our study, we have demonstrated the efficacy of simulation-based approaches and literature search strategies in enhancing the analysis of materials prior to testing, thereby reducing the need for animal testing and facilitating the development of organs, devices, and materials.

Our research highlights the potential of conductive polymers in synthesizing artificial neurons, with Polyaniline (PANI) emerging as a prime candidate due to its exceptional physical, chemical, and mechanical properties, as well as its biocompatibility. Extensive research has been conducted on the properties, preparation, and synthesis of PANI, with numerous articles discussing its biocompatibility and suitability as a conductive biopolymer.

Prospects:

In future, material analysis will pave the way for complete artificial nerve synthesis which has the ability to regenerate. More prospects include:

- ❖ Developing biocompatible materials for artificial nerves
- ❖ Creating biocomposites for enhanced nerve synthesis
- ❖ Understanding material-neural tissue interactions for new treatments
- ❖ Advancing prosthetics and implants with biocompatible materials
- ❖ Contributing to tissue engineering and biomaterials development
- ❖ Informing design of biomedical devices (pacemakers, cochlear implants, brain-computer interfaces)
- ❖ Informing regenerative medicine strategies (stem cell therapies, organ regeneration)
- ❖ Enabling personalized treatment options with tailored materials

Acknowledgment

We are thankful to the research reviewers who helped in writing and improving these guidelines.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

9. References

- [1] W. contributors, "Neuron. Wikipedia, The Free Encyclopedia.," wikipedia, [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Neuron&oldid=1239089238>. [Accessed 19 August 2024].
- [2] S. Voet, S. Srinivasan, M. Lamkanfi, and G. van Loo, "Inflammasomes in neuroinflammatory and neurodegenerative diseases," *EMBO Mol. Med.*, vol. 11, no. 6, pp. e10248, 2019.
- [3] C. Zhou, B. Liu, Y. Huang, X. Zeng, H. You, J. Li, and Y. Zhang, "The effect of four types of artificial nerve graft structures on the repair of 10-mm rat sciatic nerve gap," *J. Biomed. Mater. Res. A*, vol. 105, no. 11, pp. 3077-3085, 2017.
- [4] C. Allmeling, A. Jokuszies, K. Reimers, S. Kall, and P. M. Vogt, "Use of spider silk fibres as an innovative material in a biocompatible artificial nerve conduit," *J. Cell. Mol. Med.*, vol. 10, no. 3, pp. 770-777, 2006.
- [5] R. Kakinoki, Y. Hara, K. Yoshimoto, Y. Kaizawa, K. Hashimoto, H. Tanaka, and K. Goto, "Fabrication of artificial

- nerve conduits used in a long nerve gap: Current reviews and future studies," *Bioengineering*, vol. 11, no. 4, p. 409, 2024.
- [6] G. & C. V. Ciardelli, "Materials for peripheral nerve regeneration.,," *Macromolecular bioscience*, vol. 6(1), pp. 13-26., 2006.
- [7] C. J. A. R. K. K. S. C. C. Y. B. G. .. & V. P. M. .. Allmeling, "Spider silk fibres in artificial nerve constructs promote peripheral nerve regeneration.,," *Cell proliferation*, vol. 41(3), pp. 408-420, 2008.
- [8] C. Radtke, C. Allmeling, K. H. Waldmann, K. Reimers, K. Thies, H. C. Schenk, and P. M. Vogt, "Spider silk constructs enhance axonal regeneration and remyelination in long nerve defects in sheep.,," *PLoS one*, , vol. 6(2), p. e16990, 2011 .
- [9] Y. Zhu, C. Wu, Z. Xu, Y. Liu, H. Hu, T. Guo, and F. Li, "Light-emitting memristors for optoelectronic artificial efferent nerve.,," *Nano Letters*, , vol. 21(14), pp. 6087-6094., 2021.
- [10] X. Jiang, S. H. Lim, H. Q. Mao, and S. Y. Chew, "Current applications and future perspectives of artificial nerve conduits," *Exp. Neurol.*, vol. 223, no. 1, pp. 86-101, 2010.
- [11] B. & C. Z. Cheng, " Fabricating autologous tissue to engineer artificial nerve.,," *Microsurgery: Official Journal of the International Microsurgical Society and the European Federation of Societies for Microsurgery*, vol. 22(4), pp. 133-137., 2002.
- [12] S. P. Company, ""The purpose of gold plating for medical devices,"," Sharretts Plating Company, , 28 September 2022. [Online]. Available: <https://www.sharrettsplating.com/blog/purpose-gold-plating-medical-devices/>. [Accessed 20 august 2024].
- [13] Wikipedia contributors, " "Gold," Wikipedia," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Gold#cite_note-194. [Accessed 20 August Aug. 20, 2024.].
- [14] Wikipedia contributors, " "Platinum," Wikipedia," Wikipedia. [Online]. [Accessed 20 august 2024].
- [15] Q. Zeng, S. Yu, Z. Fan, Y. Huang, B. Song, and T. Zhou, "Nanocone-Array-Based Platinum-Iridium Oxide Neural Microelectrodes: Structure, Electrochemistry, Durability and Biocompatibility Study.,," *Nanomaterials (Basel, Switzerland)*, vol. 12(19), p. 3445 , 2022.
- [16] I. Smirnova, S. Suttipruekpong, and W. Arlt, "Feasibility study of hydrophilic and hydrophobic silica aerogels as drug delivery systems," *Journal of Non-Crystalline Solids.* , vol. 350, p. 54–60, 2004.
- [17] E. N. Zare, P. Makvandi, B. Ashtari, F. Rossi, A. Motahari, and G. Perale, "Progress in conductive polyaniline-based nanocomposites for biomedical applications: A review," *Journal of medicinal chemistry*, , vol. 63(1), pp. 1-22., 2019.
- [18] P. Humpolíček, V. Kas Č pā rkova Č, J. Pacherník, J. Stejskal, Č Bober, Z. Capakova Č, K. A. Radaskiewicz, I. Junkar and M. Lehocký Č, "The Biocompatibility of Polyaniline and Polypyrrole: A Comparative Study of Their Cytotoxicity, Embryotoxicity and Impurity Profile," *Mater. Sci. Eng., C* , vol. 91, pp. 303-310, 2018.
- [19] P. Humpolíček, K. A. Radaskiewicz, Z. Capa Č kova Č, J. Pacherník, P. Bober, V. Kasp Č rkova Č, P. Rejmontová Č, M. Lehocký Č, P. Poníž il and J. Stejskal, "Polyaniline Cryogels: Biocompatibility of Novel Conducting Macroporous Material.,," *Sci. Rep.* , vol. 8(1), pp. 1-12, 2018.
- [20] Z. Kucekova, P. Humpolicek, V. Kasparkova, T. Perecko, M. Lehocky, I. Hauerlandova, P. Saha and J. Stejskal, " Colloidal Polyaniline Dispersions: Antibacterial Activity, Cytotoxicity and Neutrophil Oxidative Burst.,," *Colloids Surf.*, vol. 116, p. 411–417, 2014.
- [21] V. Guarino, M. A. Alvarez-Perez, A. Borriello, T. Napolitano, and L. Ambrosio, "Conductive PANi/PEGDA macroporous hydrogels for nerve regeneration," *Advanced healthcare materials*, vol. 2(1), pp. 218-227, 2013.
- [22] D. Xu, L. Fan, L. Gao, Y. Xiong, Y. Wang, Q. Ye, A. Yu, H. Dai, Y. Yin, J. Cai and L. Zhang, "Micro-Nanostructured Polyaniline Assembled in Cellulose Matrix Via Interfacial Polymerization for Applications in Nerve Regeneration.,," *ACS Appl. Mater. Interfaces*, vol. 8(27), p. 17090–17097, 2016.
- [23] K. Ashtari, H. Nazari, H. Ko, P. Tebon, M. Akhshik, M. Akbari, S. N. Alhosseini, M. Mozafari, B. Mehravi, M. R. Soleimani, M. E. Warkiani, S. Ahadian and Khademhosseini, "Electrically Conductive Nanomaterials for Cardiac

Tissue Engineering," *Adv. Drug Delivery Rev*, vol. 144, p. 162–179., 2019.

- [24] S. Das, M. Sharma, D. Saharia, K. K. Sarma, E. M. Muir, and U. Bora, "Electrospun silk-polyaniline conduits for functional nerve regeneration in rat sciatic nerve injury model," *Biomedical Materials*, vol. 12(4), p. 045025, 2017.
- [25] P. Humpolicek, V. Kasparkova, P. Saha, and J. Stejskal, "Biocompatibility of polyaniline," *Synthetic metals*, Vols. 162(7-8), pp. 722-727, 2012.
- [26] Y. Arteshi, A. Aghanejad, S. Davaran, and Y. Omidi, "Biocompatible and electroconductive polyaniline-based biomaterials for electrical stimulation," *European Polymer Journal*, vol. 108, pp. 150-170, 2018.
- [27] P. R. Bidez, S. Li, A. G. MacDiarmid, E. C. Venancio, Y. Wei, and P. I. Lelkes, "Polyaniline, an electroactive polymer, supports adhesion and proliferation of cardiac myoblasts," *Journal of Biomaterials Science, Polymer Edition*, Vols. 17(1-2), pp. 199-212, 2006.
- [28] S. B. Abel, E. I. Yslas, C. R. Rivarola, and C. A. Barbero, "Synthesis of polyaniline (PANI) and functionalized polyaniline (F-PANI) nanoparticles with controlled size by solvent displacement method: Application in fluorescence detection and bacteria killing by photothermal effect," *Nanotechnology*, , vol. 29(12), p. 125604., 2018.

10. Appendix

FIGURES

1. Analysis of Gold using COMSOL

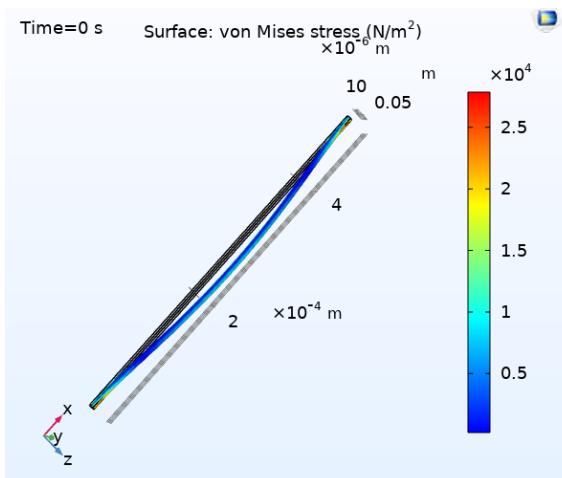


Figure 5: WHEN ALL X, Y, AND Z =10 N

2. Analysis of Iridium using COMSOL

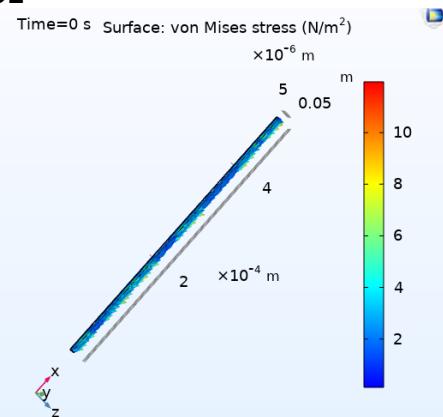


Figure 6: When x, y, and z =10n

3. Analysis of Platinum using COMSOL

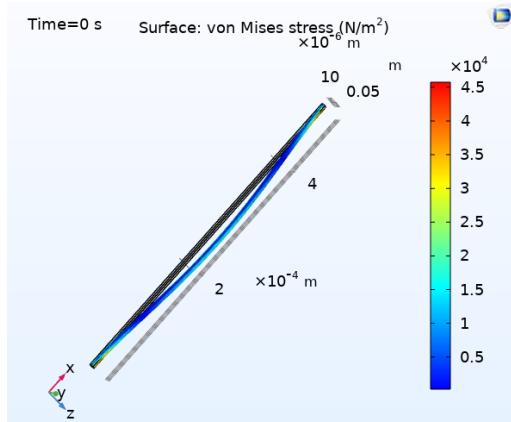


Figure 7: When X, Y, and Z = 10 N

4. Analysis of Aerogel using COMSOL

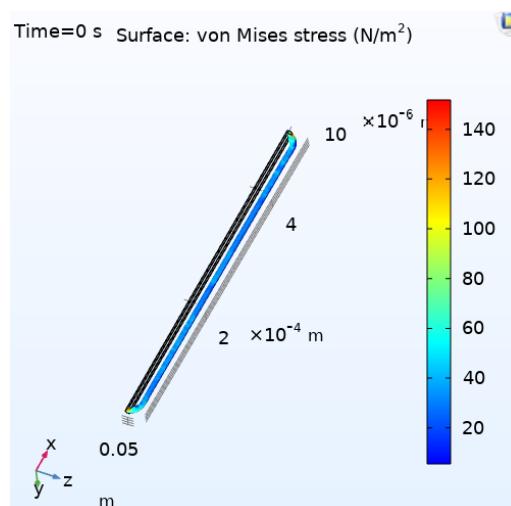


Figure 8: When x, y, and z = 10N

5. Google Scholar search results:

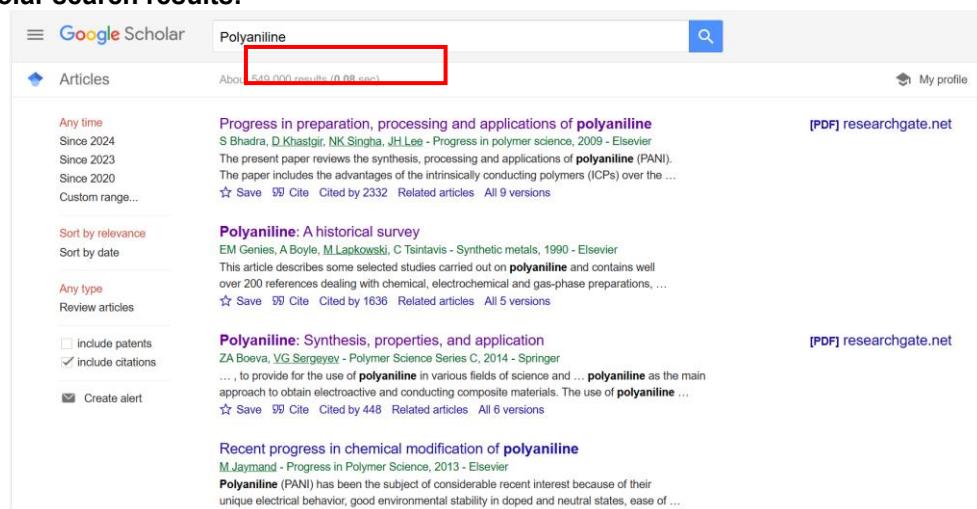


Figure 9: Google scholar search results

6. PubMed search results:



Figure 10: PubMed search result



FUSST International Conference on Computer Science and Engineering

FICSE

Research Article

Optimal Feature Selection and Multi-Class Classification of Skin Cancer Using Pre-Trained Deep Neural Networks

Veena Dillshad^a, Muhammad Nazir^b Muhammad Nouman Noor^c Muhammad Attique Khan^d

^a Department of Computer Science HITEC University Taxila, Pakistan

(veena.dillshad@hitecuni.edu.pk)

^b Department of Computer Science HITEC University Taxila, Pakistan

(muhammad.nazir@hitecuni.edu.pk)

^c Department of Computer Science Fast -NUCES Pakistan

^d Dept. of Computer Science Prince Mohammad bin Fahd University

Article Citation:

Dillshad et al. (2024). "Optimal Feature Selection and Multi-Class Classification of Skin Cancer Using Pre-Trained Deep Neural Networks". FUSST International Conference on Computer Science and Engineering



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Skin cancer is the fifth most common cancer across the globe and has a high probability of spreading to other parts of the body which makes it most dangerous and incurable once spread. So, for this disease to be curable, it is of utmost importance that it gets detected earlier. Computer-aided diagnostic systems have contributed a lot to this domain and recently deep learning algorithms have stepped forward to take the lead. Automated systems have a lot of challenges to overcome for accurate and precise diagnosis: imbalance data, low contrast images, hand-crafted features, and lesser amount of data being a few of the most critical challenges. This work proposed a fine-tuned image enhancement method utilizing mathematical operators and has used its output as an alternative way to augment dataset instead of flip and rotate. Two fine-tuned deep learning models have been utilized in this study using transfer learning and their performance has been studied by providing them balanced dataset and gradually increasing the amount of data. Deep features have been extracted and impact of optimization on the extracted features has been analyzed. ISIC2018 dataset is used for the evaluation of this work. Comparison with the state of the art is done that validates the impact of enhancement and data augmentation techniques on performance of deep learning algorithms.

Keywords: skin cancer, Lesions, Deep Learning, Image enhancement, data augmentation.

1. INTRODUCTION

The abnormal growth of skin cells, usually in sun-exposed areas such as the scalp, face, lips, ears, neck, chest, arms, and hands, as well as on women's legs, is known as skin cancer. However, it can also develop in places like your hands, the space under your finger or toenail, and your genital area that are rarely exposed to daylight [1]. Seven main classes of skin cancer exist: Bowen's disease (akiec), actinic keratoses and intraepithelial carcinoma, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions [2].

Skin cancer is the 5th most common cancer in the United States and according to estimates, 7990 (5420 men and 2570 women) deaths are expected in the year 2023. It has been observed that skin cancer is a leading cause of mortality in regions with high solar activity. By examining the statistics of skin cancer over the past two decades, an almost 5% decrease in deaths is observed from the year 2012 onwards [3]. This decline in mortality rate can be attributed to the advancements in the diagnosis and treatment of skin cancer. Dermoscopy is the conventional clinical technique used to diagnose skin cancer. Also known as "epiluminoscopy" or "epiluminescent microscopy," it describes the inspection of the skin using skin surface microscopy. The primary purpose of derm(at)oscopy is to assess pigmented skin lesions. When used by knowledgeable professionals, it can facilitate melanoma diagnosis. Dermoscopy can assist improve the

60% clinical diagnosis accuracy of melanoma made without the use of an aided eye. However, the examiner's experience has a major impact on the dermoscopy's diagnostic accuracy [4].

The application of computational intelligence has remarkably increased the diagnostic accuracy of skin cancer [5]. The survival rate of skin cancer patients depends on multiple factors. One of the prime factors in this regard is the early detection of skin cancer. Computer-aided diagnostic (CAD) systems have played a vital role in the timely detection of skin lesions. The gradual decline in the death rate of melanoma and recent studies validate the impact of computer-

Copyright

Copyright © 2024 Ujan et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

aided diagnostic systems. A standard CAD system performs the following steps: Pre-Processing, Segmentation, Feature Extraction, Feature Selection, and finally Classification.

Traditional CAD systems depend on manually extracted features for the classification of lesions. However, these hand-crafted features are inadequate to solve the challenges imposed by publicly available datasets including complicated images, interclass similarity, and intra-class differences. Recently, Deep learning has evolved as a promising tool to overcome these challenges.

Previous studies have mostly worked on the classification of skin lesions into one of the two classes either malignant or benign. This binary classification system, however, will not be effective in general clinical situations. An efficient multi-class skin cancer diagnostic system is needed for such situations. Many researchers have contributed to designing automated diagnostic tools for efficient skin lesion segmentation and classification focusing on different stages of CAD systems and incorporating different computing technologies. Along with the advancement in research and development in automated diagnostic tools, new challenges keep coming to the surface that must be addressed to increase the utility of these CAD systems. One of the biggest challenges in this regard is the availability of balanced datasets, as the publicly available datasets are highly imbalanced. Moreover, for the improved performance of deep learning models a higher number of images are required. This work proposes a framework to counter the challenges of the imbalanced dataset and analyzes the impact of an increasing number of images on the performance of deep learning models. Major contributions of this work are:

- Proposed a fine-tuned image enhancement technique based on arithmetic operators.
- Output of the enhancement step is utilized to augment the dataset instead of flipping and rotating the images.
- The augmented enhanced dataset is employed for the training of two fine-tuned deep learning (DL) models using transfer learning.
- Ant Lion Optimization algorithm is utilized for feature reduction and efficient classification.
- Analyzed the impact of increasing the number of images on the performance of data-hungry deep learning models, keeping the dataset balanced.

The rest of the paper is organized as follows: Related work is discussed in section 2, it is followed by proposed methodology in section 3, experimental results are given in section 4, and finally section 5 concludes the paper.

2.RELATED WORK

Several researchers have been actively involved in the field of automated diagnosis of skin cancer. A lot of work has been done on the pre-processing phase to aid accurate classification. Significant advancements have also been made in the accurate binary classification of skin lesions. There is still a lot to be done in the area of multiclass skin lesions classification as the six classes of skin cancer other than melanoma are proven as severe as melanoma. In addition to this, inter-class similarity and intra-class differences make this area more challenging. A major obstacle in the multiclass skin lesion classification process is the imbalanced dataset, which leads to biased outcomes. Considering all these challenges, researchers are continuously working in this domain presenting novel strategies and developing new technologies.

In [6] authors proposed a DL-based system for detection and classification of lesions. It utilizes purification and augmentation networks. ISIC 2017 and ISIC 2018 data sets are used for testing. In [7] an effective method was suggested by the researchers, which creates hybrid features by fusing data from both deep learning and manually created characteristics. These features are then fed into a multiclass SVM classifier-based decision-making model. The system evaluation makes use of the ISIC 2018 validation dataset. This work achieved a classification accuracy of 84.1%. In [8], Al-masni et al. proposed an integrated model, comprising cascading novel deep learning networks. This work proposed a new approach for the segmentation of lesion boundaries as well as the classification of skin lesions. ISIC 2017 is used for testing and evaluation. In [9], authors proposed a method to improve classification efficiency by using deep learning-based lesion segmentation. This work used U-net for segmentation and manual features are given to SVM and few other classifiers for classification. An accuracy of 85.19% is achieved by the SVM classifier. In [10], Hasan et al. presented a robust and automated skin lesion segmentation that increased the classification accuracy of the lesion. ISIC 2018 dataset is used to evaluate the proposed work. Masni et al. in [11], suggested a diagnostic paradigm that combined the segmentation of the skin lesion boundary with the classification stage of multiple skin lesions. This framework is evaluated using ISIC 2016, 2017 and 2018. In [12], Khan et al. presented automated multiclass skin lesion segmentation and classification using deep features. In [13], convolutional neural networks (ConvNet) with three architectures (InceptionV3, ResNet, and VGG19) are used to choose and train deep learning models. All these previous works have contributed to the strive for the accurate and precise diagnosis of skin cancer. Researchers have focused on the pre-processing step and segmentation step to improve classification results. Most of the previous work has focused on data augmentation by only deploying flips and rotations on the dataset. Our proposed work has contributed to image enhancement and has utilized these enhanced images to augment the dataset. This step aims to increase the classification accuracy without adding the additional steps of image segmentation.

3.PROPOSED METHODOLOGY

This section elaborates on the mathematical and theoretical details of the proposed work for multi-class skin lesion classification. Figure 1 illustrates the many stages of the proposed approach, which include data augmentation and enhancement before employing transfer learning to train two deep learning models on the enhanced dataset. This led to the extraction of features from these fine-tuned deep-learning models. These extracted features are then optimized

Dillshad et al. "Optimal Feature Selection and Multi-Class Classification of Skin Cancer" using an optimization technique to remove the irrelevant features and reduce the classification time. The last step is the classification of these selected features using machine learning classifiers.

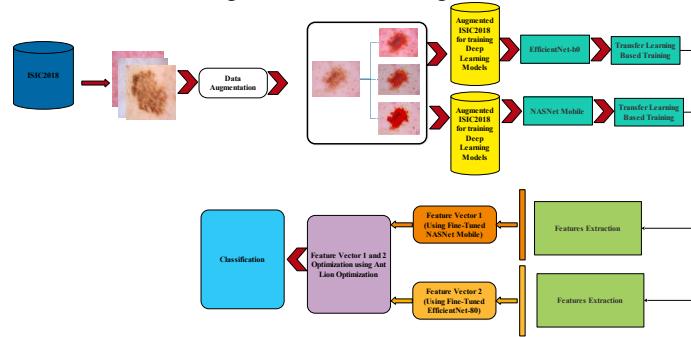


Figure 1: Proposed Framework

3.1. Dataset

The suggested study is evaluated using the publicly available ISIC 2018 dataset. The International Skin Imaging Collaboration (ISIC) released the ISIC 2018 dataset, a sizable collection of dermoscopy images [14]. The following is the distribution of images in each class: 327 images in Actinic keratoses (AK), 514 images in Basal cell carcinoma (bcc), 1099 images in Benign keratosis-like lesions (bkl), 115 images in Dermatofibroma (df), 1113 images in Melanoma (mel), 6705 images in Melanocytic nevi (nv), 142 images in Vascular lesions (vasc). The stark difference in the number of photographs in each class makes the imbalanced dataset issue clear and causes skewed findings.

Figure 2: display sample images taken from the dataset.

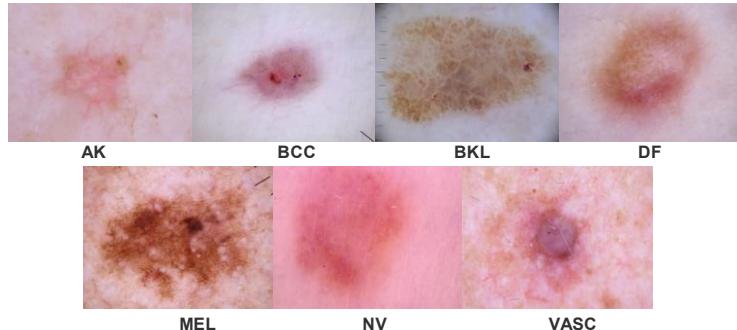


Figure 2: Sample images from the ISIC 2018 dataset

3.2. Data augmentation

This section describes the mathematical operations used for the enhancement of images. The output of this enhancement step is further utilized to augment the dataset as well. Saving these enhanced outputs to the dataset caters to the problem of an imbalanced dataset and will improve the quality of the image as well. The proposed enhancement method is based on the mathematical operations given below. The impact of these operations on the images is also highlighted in the figures. Images from the original dataset are preprocessed using equation (1) to enhance the lesion and make the affected area differentiable from healthy skin. Where $\mathcal{B}'(x,y)$ represents the output image, $\mathcal{B}(x,y)$ denotes the input image, $\mathcal{B}c(x,y)$ denotes the image compliment and α is the amount of illumination added to brighten the image. α is set to 50 for this work. The impact of this mathematical operation on the original image is shown in Figure 3.

$$\mathcal{B}'(x,y) = (\mathcal{B}(x,y) - \mathcal{B}c(x,y)) + \alpha \quad (1)$$

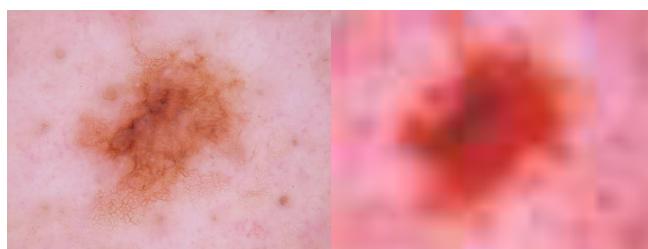


Figure 3: Sample images from the ISIC 2018 dataset

The second mathematical operation designed to enhance images is shown in equation (2). In this operation the intensity of original image is enhanced by the factor α_1 , then the intensity of image compliment is multiplied by the factor α_2 . Finally, this image complement is subtracted from the intensity amplified image to get the enhanced output image. α_1, α_2 are carefully selected to be 0.5 and 0.75 respectively. The impact of this operation on the original

Dillshad et al. "Optimal Feature Selection and Multi-Class Classification of Skin Cancer" image is shown in figure 4, which clearly distinguishes the lesion from healthy skin in comparison to the original poor contrast image.

$$\mathbb{B}'(x,y) = (\mathbb{B}(x,y) + \alpha_1 * (\mathbb{B}(x,y))) - ((\mathbb{B}_c(x,y)) * \alpha_2) \quad (2)$$

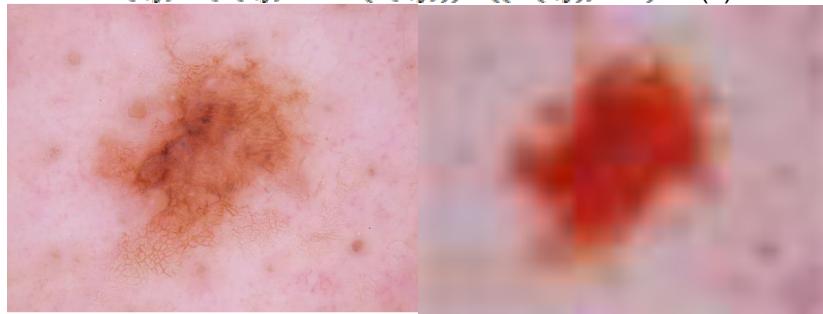


Figure 2: Sample images from the ISIC 2018 dataset

The third mathematical operation is shown in equation (3). In this operation, a two-step procedure is performed. The first image is scaled down with the factor of α_1 and then this image is multiplied with the original image. These image arithmetic operations caused hazy lesion images to stand out from the background as shown in Figure 5.

$$\mathbb{B}''(x,y) = (\mathbb{B}(x,y) * (\alpha_1 * (\mathbb{B}(x,y)))) \quad (3)$$

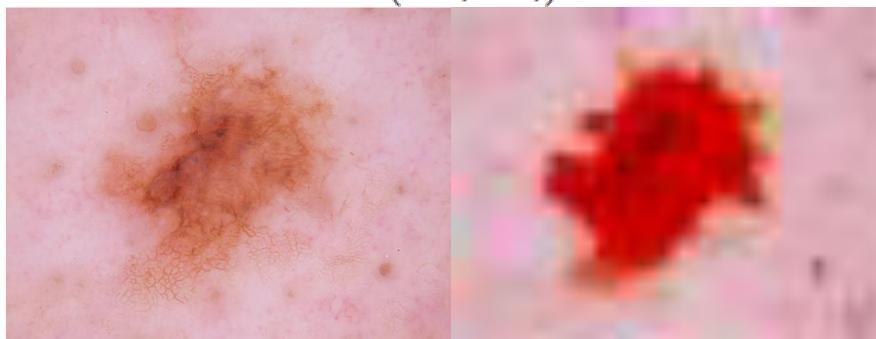


Figure 2: Sample images from the ISIC 2018 dataset

Output of each of these operations is saved for augmenting the dataset. Total number of images after data augmentation are given in Table 1.

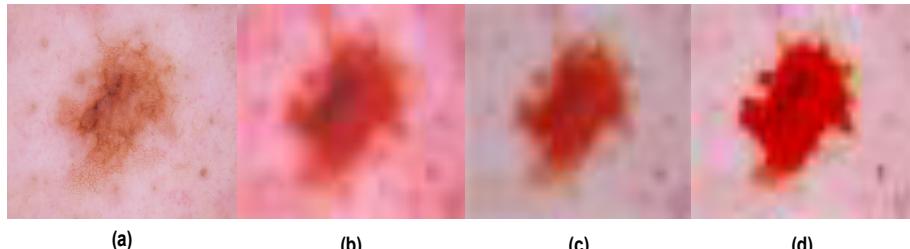


Figure 2: Sample images from the ISIC 2018 dataset

Table 3.1: Image Distribution in the original ISIC 2018 Dataset and augmented ISIC 2018 Dataset

Class/ Dataset	ISIC 2018	Augmented ISIC 2018 (1000 images in each class)	Augmented ISIC 2018 (2000 images in each class)	Augmented ISIC 2018 (3000 images in each class)
AK	327	1008	2016	3002
BCC	514	1028	2056	3086
BKL	1099	1000	2000	3000
DF	115	1035	2070	3022
MEL	1113	1000	2000	3000
NV	6705	1000	2000	3000
VASC	142	1136	2272	3024

3.3. Pre-trained Deep Learning Models

In the proposed framework, two DL models have been utilized, NASNet Mobile and EfficientNet B0. These two models are selected as they light weight and select hyper parameters optimally instead of manual selection, making them the best choice for efficient clinical diagnosis. EfficientNet is a family model whose basic building block is developed by the NAS (Neural Architectural Search). NAS is a framework that searches for the best architecture with the help of controller, instead of manually setting the hyper parameters, under certain conditions. The main building block of the efficient NET b0 network is the mobile inverted bottleneck (MBConv) layer. The performance of EfficientNet is far better than all the previous models [15].

The NASNet (Neural Architecture Search) Mobile is the second model used in the proposed framework. It automates the process of network architecture selection. NASNet searches for the best algorithm to achieve maximum performance on a particular task. The main objective of NASNet is to find the most suitable combinations from a set of operations with the help of controller RNN to build a block with the best performance instead of designing the block using manual decisions [16].

These two models are trained on an augmented ISIC 2018 dataset using deep transfer learning. Originally these models were trained on ImageNet Dataset having 1000 classes at output layer. After fine tuning these models, knowledge is transferred. In the later steps, these fine-tuned models are trained on an augmented ISIC 2018 dataset with seven classes at the output layer.

3.4. Feature Extraction

Training of these deep learning models lead to two modified models, Modified Efficient Net b0 and Modified NASNet Mobile. The Global Average Pooling (GAP) layer is chosen and activated for Efficient Net B0 that has been tuned. As an activation function, the entropy loss function is used. 1280 features are extracted for each image on this layer. Similar to this, the GAP layer is chosen and activated for fine-tuned NASNet Mobile. As an activation function, the entropy loss function generated 1056 features for each image. We employed several hyperparameters throughout the training phase, including a learning rate of 0.0001, mini-batch size of 32, stochastic gradient descent as the optimization approach, and momentum of 0.7. The following step involves optimizing the extracted features individually to study their impact on classification accuracy.

3.5. Feature Optimization

The extracted features are optimized using nature inspired Ant Lion Optimization algorithm [17]. This program simulates the antlions' natural behavior to hunt ants. The global optimum for optimization problems is approximated by the triplet function known as the ALO method, which looks like this: $ALO(A, B, C)$

A: a formula that yields the initial random solutions, *B*: population that *A* provides, and when the end condition is met, *C* returns true.

```

Initialize the first population of ants and antlions randomly
Calculate the fitness of ants and antlions
Find the best antlions and assume it as the elite (determined optimum)
while the end criterion is not satisfied
    for every ant
        Select an antlion using Roulette wheel
        Update c and d using equations
        Create a random walk and normalize it

        Update the position of ant
    end for
    Calculate the fitness of all ants
    Replace an antlion with its corresponding ant it if becomes fitter
    Update elite if an antlion becomes fitter than the elite
end while
Return elite

```

Figure 5: Pseudocode for Ant Lion Optimization Algorithm [17]

4. EXPERIMENTAL RESULTS

This section elaborates on the outcomes of the proposed framework. The augmented ISIC 2018 dataset is utilized for the assessment of this work. The training and testing ratio for this work is taken to be 50:50. 10-fold cross validation is used for testing results. For the classification, several machine learning classifiers are utilized namely, Linear SVM, Medium Gaussian SVM, Quadratic SVM, Coarse Gaussian SVM, Fine KNN, Coarse KNN, medium KNN, Cosine KNN, Cubic KNN, and Weighted KNN. Accuracy is taken as a performance metric and is computed for each classifier. MATLAB2022b is used to simulate the full structure on a computer with 16 GB of RAM and an 8 GB graphics card.

Three sets of experiments have been conducted to study the effect of an increasing number of images on the performance of fine-tuned DL models. In the first experiment the proposed framework is implemented using an augmented balanced ISIC 2018 dataset with 1000 images in each class results of this experiment are displayed in Table 2. It can be seen that Cubic SVM attained maximum accuracy in the experiments performed using both NASNet Mobile and EfficientNet b0 with 78.3% classification accuracy. These results also show the classification accuracy of

Dillshad et al. "Optimal Feature Selection and Multi-Class Classification of Skin Cancer"
the optimized features, where the Ant lion optimization algorithm has reduced the feature set significantly while preserving the classification accuracy.

Table 2: Classification results of Experiment 1

Classifier/ Model	NASNet Mobile		EfficientNet B0	
	Accuracy % (before optimization) 1056 features	Accuracy % (After optimization) 695 features	Accuracy % (before optimization) 1280 features	Accuracy % (After optimization) 662 features
Linear SVM	73.9	74.7	76.6	73.5
Quadratic SVM	74.3	75.5	77.6	76.1
Cubic SVM	75.0	75.7	78.3	76.5
Fine Gaussian SVM	20.5	20.0	17.1	17.1
Medium Gaussian SVM	74.4	75.5	77.3	75.5
Coarse Gaussian SVM	72.5	73.4	71.8	71.3
Fine KNN	69.7	69.2	71.6	71.0
Medium KNN	72.0	73.3	70.6	70.9
Coarse KNN	70.5	71.6	67.9	67.2
Cosine KNN	73.2	73.2	73.3	73.1
Cubic KNN	71.6	72.8	70.5	70.2
Weighted KNN	73.3	73.4	72.6	72.1

These results can be confirmed by the confusion matrices of Cubic SVM classifier using Fine-tune NASNet Mobile and EfficientNet b0 in Figure 8 and 9 respectively.

Model 2.3								
True Class	AK	517	44	39	8	24	20	1
	BCC	46	387	19	6	23	20	13
	BKL	20	13	384	4	82	42	4
	DF	22	14	10	359	15	27	13
	MEL	15	8	66	3	377	77	10
	NV	1	5	27	3	77	398	11
	VASC	1	15	4	14	19	70	445
Predicted Class								

Figure 6: Confusion Matrix of Cubic SVM using Fine-tuned NASNet Mobile

Model 2.3								
True Class	AK	529	51	33	11	22	5	2
	BCC	56	400	31	7	12	7	1
	BKL	17	14	390	2	87	38	1
	DF	31	24	8	355	12	9	21
	MEL	15	10	84	1	392	52	2
	NV	5	7	33	2	66	404	5
	VASC	5	9	2	24		6	522
Predicted Class								

Figure 7: Confusion Matrix of Cubic SVM using Fine-tuned EfficientNet b0

The second experiment follows the same steps but with 2000 images in each class of balanced dataset. Table 3 presents the results of this experiment. A remarkable increase in the classification accuracy can be observed in the second experiment that validates the fact that deep learning algorithms perform significantly well with large number of training images. Cubic SVM has outperformed rest of the classifiers using both NASNet Mobile and EfficientNet b0 by achieving 84.1% highest classification accuracy which is significantly high as compared to the experiment 1 i.e. 78.3%. Same trend can be observed in the results of classification after optimization, where the feature selection has been performed well while preserving the classification accuracy.

Classifier/Model

	NASNet Mobile		EfficientNet B0	
	Accuracy % (before optimization) 1056 features	Accuracy % (After optimization) 559 features	Accuracy % (before optimization) 1280 features	Accuracy % (After optimization) 791 features
Linear SVM	81.9	81.2	79.0	78.2
Quadratic SVM	83.2	82.7	81.8	80.0
Cubic SVM	84.1	83.5	83.0	81.2
Fine Gaussian SVM	20.3	20.2	16.7	16.8
Medium Gaussian SVM	82.2	81.9	80.9	80.0
Coarse Gaussian SVM	79.6	78.3	75.6	74.0
Fine KNN	76.8	76.9	79.1	78.9
Medium KNN	78.8	78.4	76.4	76.0
Coarse KNN	76.7	75.9	70.9	70.0
Cosine KNN	79.5	79.4	78.6	77.4
Cubic KNN	78.4	77.8	76.1	75.4
Weighted KNN	79.5	79.2	78.2	78.1

These results can be verified by the confusion matrices of Cubic SVM classifier using Fine-tune NASNet Mobile and EfficientNet b0 in Figure 10 and 11 respectively.

Model 2.3							
True Class	AK	48	57	6	19	10	11
	BCC	50	882	28	7	14	20
	BKL	33	25	828	5	108	94
	DF	17	13	8	800	6	11
	MEL	21	12	110	7	803	157
	NV	2	15	59	4	105	1062
	VASC	6	17	1	20	1	10
Predicted Class							

Figure 8: Confusion Matrix of Cubic SVM using Fine-tuned NASNet Mobile

Model 2.3							
True Class	AK	81	55	22	41	9	3
	BCC	66	871	48	13	13	15
	BKL	49	39	811	5	122	71
	DF	41	32	11	772	13	15
	MEL	25	23	123	4	811	124
	NV	6	17	51	2	89	1091
	VASC	5	13	11	27	3	1076
Predicted Class							

Figure 9: Confusion Matrix of Cubic SVM using Fine-tuned EfficientNet b0

In the third experiment, the number of images in each class of balanced dataset further increased to 3000. The impact of this increase is clearly shown in results as displayed in Table 4. Increase in the classification accuracy can be observed in third experiment that again validates the fact that training data quantity affects the performance of a data-hungry algorithms. Cubic SVM has achieved highest classification accuracy of 86.2% which is a remarkable increase as compared to the experiment 1 i.e. 78.3%. As far as classification results after optimization are concerned, it helped in feature reduction and accuracy is retained.

Classifier/ Model

	NASNet Mobile		EfficientNet B0	
	Accuracy % (before optimization) 1056 features	Accuracy % (After optimization) 833 features	Accuracy % (before optimization) 1280 features	Accuracy % (After optimization) 1237 features
<i>Linear SVM</i>	84.3%	84.2%	77.4%	77.5%
<i>Quadratic SVM</i>	85.4%	85.5%	80.3%	80.7%
Cubic SVM	86.2%	86.1%	81.7%	82.1%
<i>Fine Gaussian SVM</i>	27.2%	26.3%	17.5%	17.5%
<i>Medium Gaussian SVM</i>	85.1%	85.2%	80.0%	80.0%
<i>Coarse Gaussian SVM</i>	82.7%	82.6%	74.3%	74.3%
<i>Fine KNN</i>	85.1%	84.7%	81.7%	82.5%
<i>Medium KNN</i>	82.8%	82.8%	77.2%	77.5%
<i>Coarse KNN</i>	80.5%	81.0%	69.2%	69.4%
<i>Cosine KNN</i>	83.6%	82.9%	78.6%	79.3%
<i>Cubic KNN</i>	82.0%	81.9%	76.5%	77.1%
<i>Weighted KNN</i>	84.1%	84.2%	78.9%	80.0%

These results can be verified by the confusion matrices of Cubic SVM classifier using Fine-tune NASNet Mobile in Figure 12.

Model 2.3								
True Class	AK	1321	53	65	17	29	5	11
	BCC	37	1398	42	9	25	7	25
	BKL	57	34	1199	13	173	60	8
	DF	18	11	10	1423	7	3	39
	MEL	25	21	139	12	1165	141	19
	NV	3	17	58	2	151	1252	19
	VASC		15	7	69	10	2	1409
	Predicted Class							

Figure 10: Confusion Matrix of Cubic SVM using Fine-tuned NASNet Mobile

The experimental results of the proposed work strengthen the fact that performance of deep learning algorithms improve with the increase in amount of training data. The impact of increased dataset on performance is visually represented in figure 13. To validate the impact of proposed image enhancement operations and dataset augmentation, the comparison of this work with the state of the art is given in Table 5. It can be observed that the proposed work has outperformed most of the previous works without additional steps like segmentation.

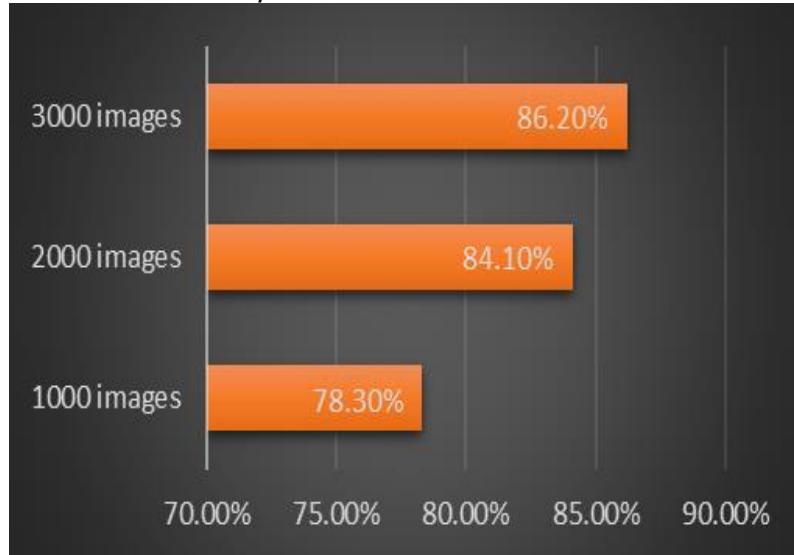


Figure 11: Impact of Size of Dataset on Classification Accuracy

Table 5: Comparison with the State-Of-The-Art Techniques

Sr#	State Of the art	Classification Accuracy
1.	[18]	75.4%
2.	[7]	84.1%
3.	[19]	76.0%
4.	[20]	86.2%
5.	[21]	83.0%
6.	[22]	73.03%
7.	Proposed Framework	86.2%

5. CONCLUSIONS

This research proposes a multiclass skin lesion categorization framework based on deep learning focusing on image enhancement and data augmentation. This study also studied the impact of variations in a number of images on the performance of DL models. In the initial steps of framework, images are enhanced using mathematical operations, output of this step is utilized to augment dataset. Two pre-trained deep learning lightweight models are adjusted and trained using enhanced datasets in the following stage. After extracting deep features from fine-tuned models these features are optimized using Ant Lion Optimization technique which provides the relevant features to improve efficiency of the proposed work. Finally, several machine learning classifiers produce the classification results. This work has provided improved classification accuracy on ISIC 2018 dataset by solving the imbalanced dataset problem. In future, these results can be further improved by adding a fusion step in the framework. Moreover, an improved optimization algorithm can be used that not only reduces the feature set but also improves accuracy.

6. REFERENCES

- [1] Mayo Clinic. Skin cancer: Symptoms and causes. Mayo Clinic. Retrieved June 19, 2024, from <https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605>
- [2] Cancer Treatment Centers of America. **Types of skin cancer**. Cancer Treatment Centers of America. Retrieved June 19, 2024, from <https://www.cancercenter.com/cancer-types/skin-cancer/types>
- [3] American Cancer Society. *Cancer Facts & Figures 2021*. Atlanta: American Cancer Society; 2021.
- [4] Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions--a valuable tool for early diagnosis of melanoma. *Lancet Oncol*. 2001 Jul;2(7):443-9. doi: 10.1016/s1470-2045(00)00422-8. PMID: 11905739.."
- [5] Ammara Masood, Adel Ali Al-Jumaily, "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms", International Journal of Biomedical Imaging, vol. 2013, Article ID 323268, 22 pages, 2013. <https://doi.org/10.1155/2013/323268>".
- [6] Khan, Muhammad Attique, Muhammad Imran Sharif, Mudassar Raza, Almas Anjum, Tanzila Saba, and Shafqat Ali Shad. "Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection." *Expert Systems* 39, no. 7 (2022): e12497.
- [7] Hardie, Russell C., Redha Ali, Manawaduge Supun De Silva, and Temesguen Messay Kebede. "Skin lesion segmentation and classification for ISIC 2018 using traditional classifiers with hand-crafted features." *arXiv preprint arXiv:1807.07001* (2018).

- Dillshad et al. "Optimal Feature Selection and Multi-Class Classification of Skin Cancer"
- [8] Al-Masni, Mohammed A., Mugahed A. Al-Antari, Hye Min Park, Na Hyeon Park, and Tae-Seong Kim. "A deep learning model integrating FrCN and residual convolutional networks for skin lesion segmentation and classification." In *2019 IEEE Eurasia conference on biomedical engineering, healthcare and sustainability (ECBIOS)*, pp. 95-98. IEEE, 2019.
- [9] Seeja, R. D., and A. Suresh. "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM)." *Asian Pacific journal of cancer prevention: APJCP* 20, no. 5 (2019): 1555.]
- [10] Hasan, Sohaib Najat, Murat Gezer, Raghad Abdulaali Azeez, and Sevinç Gülsen. "Skin lesion segmentation by using deep learning techniques." In *2019 Medical Technologies Congress (TIPTEKNO)*, pp. 1-4. IEEE, 2019
- [11] Al-Masni, Mohammed A., Dong-Hyun Kim, and Tae-Seong Kim. "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification." *Computer methods and programs in biomedicine* 190 (2020): 105351.
- [12] Khan, Muhammad Attique, Muhammad Sharif, Tallha Akram, Robertas Damaševičius, and Rytis Maskeliūnas. "Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization." *Diagnostics* 11, no. 5 (2021): 811.
- [13] Mijwil, Maad M. "Skin cancer disease images classification using deep learning solutions." *Multimedia Tools and Applications* 80, no. 17 (2021): 26255-26271.
- [14] Codella, Noel, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba et al. "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)." *arXiv preprint arXiv:1902.03368* (2019).
- [15] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In *International conference on machine learning*, pp. 6105-6114. PMLR, 2019.
- [16] Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018
- [17] Mirjalili, Seyedali. "The ant lion optimizer." *Advances in engineering software* 83 (2015): 80-98.
- [18] Bisla, Devansh, Anna Choromanska, Jennifer A. Stein, David Polksy, and Russell Berman. "Skin lesion segmentation and classification with deep learning system." *arXiv preprint arXiv:1902.06061* (2019): 1-6.
- [19] Milton, Md Ashraful Alam. "Automated skin lesion classification using ensemble of deep neural networks in isic 2018: Skin lesion analysis towards melanoma detection challenge." *arXiv preprint arXiv:1901.10802* (2019).
- [20] Mahbod, Amirreza, Gerald Schaefer, Chunliang Wang, Georg Dorffner, Rupert Ecker, and Isabella Ellinger. "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification." *Computer methods and programs in biomedicine* 193 (2020): 105475.
- [21] Budhiman, Arief, Suyanto Suyanto, and Anditya Arifianto. "Melanoma cancer classification using resnet with data augmentation." In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 17-20. IEEE, 2019.
- [22] Hardie, Russell C., Redha Ali, Manawaduge Supun De Silva, and Temesguen Messay Kebede. "Skin lesion segmentation and classification for ISIC 2018 using traditional classifiers with hand-crafted features." *arXiv preprint arXiv:1807.07001* (2018).

Article Citation:
Author1_LastName et al.
(2024). "A Short, Easy-to-
Comprehend Title
Representing Your Research
Idea". *FUSST International
Conference on Computer
Science and Engineering*



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024
Author1_LastName et al.



Published by
Foundation University
Islamabad.

Web: <https://fui.edu.pk/>

detection techniques. Not only that, PV modules are thinning such that they can easily be broken down by environmental degradation which makes them both less performant and safe [3].

Infrared thermography (IRT) and electroluminescence (EL) have recently emerged as defect detection techniques for photovoltaic systems that are non-destructive and effective approaches to identifying faults in PV systems [4]. Through periodic maintenance, by utilizing these methods during the inspections, it will help in early detection of the cells under defects and replacement shall be done on time to maintain both power continuity and efficiency. Since AI is improving conventional MPPT, power prediction techniques, fault detection can enhance the system performance further [5]. Surface soiling, cracks, and discoloration are common defects that, if not detected earlier, may lead to performance reduction of the system and even PID-related issues like hot spots.

SolarAI: Solar-Panel Optimization & Defect Resolution using CNN

Muhammad Haseeb Mirza ^{a,*}, Fahad Burhan Ahmad ^a, Muhammad Habib ^{a,*}, Ruqia Bibi ^a, Rehan Mehmood Yousaf ^a, Iqra Waseem ^a, Muhammad Farzan ^a, Noman Khan ^a

^a University Institute of Information Technology, PMAS-Arid Agriculture University Rawalpindi, Pakistan.

* Corresponding author: muhammad.habib@uaar.edu.pk

Abstract:

This research explores the use of numerous concepts of intelligent or artificially generated approaches, in order to improve the nature of identifying faults on photovoltaic photovoltaic (PV) systems, which is becoming relevant due to the need for efficient solutions of solar energy use. Existing solutions such as Ground Fault Detection Interrupters (GFDI) and Overcurrent Protection Devices (OCPD) are partially effective in some situations and overly ineffective in others. Some of the prominent defects, for example microcracks blocking grid lines, surface dust, and electrical arcing are detected using CNN, SVM and colour/ O morphology semantic segmentation methods which increase detection and enhance operation a efficiency. This work proposes a new two stage autonomous recognitional system based on electroluminescence and infrared thermography for fault identification. It is shown in the experiments, that the application of the proposed AI approaches for the solar power systems either reduces the costs that already existed or increases the efficiency of the proposed so far systems. This work presents the situation which obliges to continue even more intense developments in the field of renewable energy on the increase solar energy consumption and factors that may help in future optimization of defect detection in different implementations of PV systems. Keywords: Photovoltaics; fault detection, Deep Learning, CNN, Solar Panel Defect, artificial intelligence, maintenance; cleaner energy. In this study we demonstrate the benefits, in terms and characteristics of augmentation of the PV systems quality assurance of occurrence of failure of any component or system of component, with some combinations of intelligent, or artificially generated, approaches whereby multiple approaches are used.

Keywords: Photovoltaics; fault detection, Deep Learning, CNN, Solar Panel Defect, Artificial Intelligence, Maintenance; Renewable Energy.

1. Introduction

The increasing dependency on energy consumption of modern society created by means of developed technologies comes with adverse effects as well. This increase has resulted in elevated CO₂ emissions, exacerbating environmental challenges [1]. As a countermeasure, solar energy has gained widespread attention as a viable alternative. By 2019, the global installed solar PV capacity had reached 627 GWDC, indicating considerable growth and underscoring solar energy's growing importance in the energy sector [2]. However, the reliability and efficiency of solar panels remain critical concerns, as external environmental factors directly influence their performance. Therefore, fault detection is essential to prevent power outages and ensure system reliability. Various methods, including sensor signals and environmental data, are used to detect faults in PV systems. With the projected expansion of solar PV capacity by 2030, the global shift toward solar energy aims not only to reduce dependence on fossil fuels but also to extend electricity access to isolated regions. However, traditional fault detection methods such as Ground Fault Detection Interrupters (GFDI) and Overcurrent Protection Devices (OCPD) often fail under specific conditions, highlighting the need for advanced, AI-based fault

This research emphasizes the value of automated fault detection for enhancing operational P. AI efficiency of PV systems, with a keen interest in deep-learning-based solar panel defect detection techniques. The paper therefore reviews AI-based approaches with special reference to neural networks and metaheuristic algorithms for their possible potential in improving crack identification along with other fatal defects on the solar panel surfaces. It also discusses ensemble learning and deep learning models in detecting cracks and other fatal defects. Although, despite much progress made by researchers, up to date there is no universal method for fault detecting all attendant PV systems.

The current research proposes a two-stage automatic recognition system using EL imaging for which the approach can explicitly detect both defective and non-defective cells even in low computational devices. The experimental results on a benchmark dataset of EL images from 2,624 solar cells prove that this method leads to higher accuracy in comparison with traditional fault detection techniques [7]. The paper stresses the importance of accurate defect detection for achieving the highest possible performance of PV systems and gives prospects for future research to further improve power efficiency via enhanced fault detection techniques.

2. Related Work

Various innovative techniques have been used in the defect detection of solar panels. The approach by Schuss et al. (2021) suggested thermal imaging as a new technique, with particularly good results in outdoor applications [8]. Time-Resolved Thermography and Synchronized Thermography have been used to detect the presence of defects in solar panels, as remarked by another research work [9]. In this case, deep learning methods like AlexNet using deep CNNs and transfer learning were applied by Zyout and Oatawneh (2020) and Li et al. (2019) for automated surface inspection of solar panels to apply them since that is also another recommendation from where this information was obtained [10]. Guo and Cai (2020) have introduced a thermography-based technique using an infrared camera that captures images combined with step-heating halogen lights to enhance defects' detectability in solar cells [11]. Natarajan et al. (2020), earlier presented a machine learning approach for classifying photovoltaic cells based on heat generation patterns with great efficiency: 97% accuracy rate [12].

Mal-making Solar panels can cause unscheduled power interruptions which give rise to increased expenses and poor dependability of PV systems. Aging systems need to be put into more consideration due to the risks associated with the failure of the panels, where there is a risk of electrical fires and other risks which are posed by the malfunctioning panels [13]. Non-destructive methods like laser doppler vibrometry, thermography, and electroluminescence have been utilized to determine the cracks in the solar cells but not so many authors have utilized the algorithm of the jumping gene genetic algorithm to study the dynamic illuminations and current-voltage characteristics including current to voltage founded materials [14]. There are also approaches based on images that have been used for fault detection as they are more efficient and reliable than sensor-based approaches. For instance, Shihavuddin and his colleagues employed deep Convolution Neural Networks for image classification with the Froggles architecture in this case bolstered by the layers of CNN trained to evaluate the deformation of 3,336 thermal images of PV cells [15]. Likewise, Pierdicca and colleagues focused on the use of visual pattern recognition for the detection of faults in PV systems [16]. The networks including EfficientDet, YOLO-v3, YOLO-v4, and YOLO-v5 which are built for object recognition and detection by the CNN mechanism helped identify damages in PV cells and wind turbines with sensitivity average being 0.79 [17].

Besides these approaches, El-Baby et al. [18] discuss several aspects concerning issues embedded in PV systems. Eltuhamy et al. introduced a neuro-fuzzy adaptive inference system to detect and classify defects in Copper Indium Gallium Selenide (CIGS) thin-film PV modules [19]. Memon et al. proposed a CNN-based method for identifying defects in solar panels, which was trained on historical data [20]. Chen and colleagues developed a sequential fault detection method using autoregressive models and Generalized Likelihood Ratio (GLLR) tests, with an emphasis on flexibility and fast fault identification [21]. Ramirez et al. introduced a cost-effective condition monitoring system that combines image processing with radiometric sensors for detecting faults in PV systems [22]. To enhance spatial analysis of defect hotspots, Tang et al. proposed a two-layer detection method that utilizes an orthotile-based georeferenced spatial heat map [23].

3. Proposed Methodology

Accurately diagnosing defective solar cells relies heavily on analyzing the distinctive features of cell images. Clearly defining these features is crucial to improving accuracy and minimizing classification errors in fault detection. As shown in Fig. 1, solar cell images reveal various structural characteristics that can indicate whether a cell is defective. However, the similarities among these images may lead to confusion, decreasing detection precision. To address this challenge, our proposed method focuses on identifying the specific features that differentiate defective cells from non-defective ones within clusters of similar cells. This approach aims to enhance detection accuracy, particularly in the face of variable cell morphologies [23].



Figure 1: First row shows defective and second row shows non-defective solar panels

Data Collection and Preparation

As regards the methodology, the first step involves an exhaustive collection of the image data of solar cells, then, the second step will bias annotation each image for defects as well shown in Fig 2. These topical annotations constitute the ground truth for the training and testing of the deep learning models. The dataset contains a rich mix of defects such as micro-cracks, inactive regions, surface soiling, among other factors, so that there is a good possibility that the models will learn to segment and classify different types of defects in the fabricated devices.

Deep Learning Models for crack segmentation

For crack segmentation four deep learning models were chosen, Model A, Model B, Model C and Model D; Model A, Model B, Model C, and Model D. Each model was subjected to training using the same annotated dataset and the corresponding hyper-parameters in order to conserve uniformity. In addition, the models were tested against key performance metrics including accuracy, precision, recall, and F1 score. Attempted in the current work was comparison of the various important thin layer models performance quite reliably optimizing these for computational speed and reliability against differing defect types.

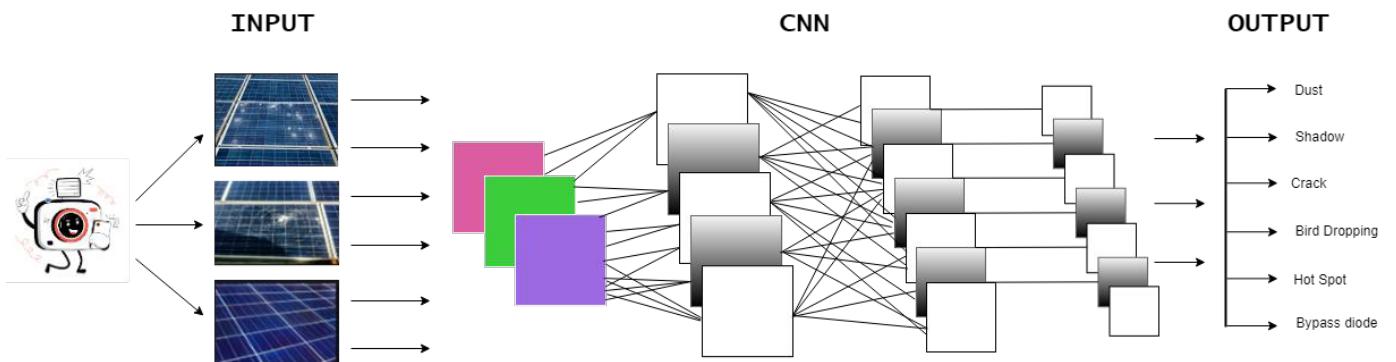


Figure 2: Proposed Model diagram

Ensemble Learning for Improved Accuracy

Enhancing the specificity and generalization of credits was performed by way of average ensemble learning of the four deep learning models outputs. This method is more reliable as it exploits the merits of all the models and therefore, enhances defect detection system. We average the predictions of all models because isolating failures in one or more models leads to a compromised detection process.

Impact of Fracture Types on Power Efficiency

For the last step, a particular investigation was undertaken, directed to the evaluation of the influence of individual cracks on the power efficiency of solar cells. This study examines the relationship between defect type and defect magnitude and the reduction in power output associated with that defect, thus demonstrating why count and measure the faults and the importance of early intervention in maintaining the efficiency of the system.

4. Experiments

Experimental Setup

The dataset employed in this study was sourced from the open access DuraMAT DataHub (Electroluminescence Image Analysis – Datasets, DuraMAT [Online]) [25] that has 576 electroluminescence (EL) images of solar panels. Additionally, 1837 images of solar cells were retrieved from one of the Github repositories [26]. These images were manually annotated using Computer Vision Annotation Tool (CVAT) and ground truth masks were produced. To

address this problem in terms of deterioration of the performance of PV cells, five classes were established for deep fracture and retaining dendritic crack in a few directions whilst including horizontal and vertical as well as angular cracks including the solar cell breakdowns. In order to reduce noise and enhance the segmentation accuracy, Busbars were also able to hitch up.

The 150x150 RGB images were changed to 128x128 monochrome as a measure of saving the processing burden while not losing any relevant detail. The dataset was divided into three predominant parts namely training, validation and testing leave 70, 20, and 10 percent respectively, thus providing a fair assessment of the model.

Convolutional Neural Networks for Defect Detection

Due to the superiority of CNNs in image-tasks, they were used in detecting defects. These include imaged based grids designed models which apply hierarchical pattern recognition for images through convolutional, pooling and fully connected layers which are grouped themselves as deep learning models. Particularly, CNN combines the advantages of both biological visual system and machine learning to learn spatial features, making it suitable for solar cell structural defect detection from images [27].

The layers which formed the architecture of the CNN model in this research work were as follows:

Convolutional (CONV): Extracting features through applying filters across the input image.

Activation (RELU): Leads non-linearity in model.

Pooling (POOL): Lessens spatial dimensions while retaining important features.

Fully Connected (FC): Maps learned features to final output.

Batch Normalization (BN): Stabilizes and accelerates training by normalizing input values.

Dropout (DO): Prevents overfitting by randomly deactivating neurons during training.

The batch normalization formula used in training is as follows:

$$\hat{X}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \quad (1)$$

where μ_β and σ_β are the mean and variance calculated over each mini-batch, and ϵ is a small constant to prevent division by zero.

Batch normalization is applied before the activation function to stabilize and improve the learning process by ensuring the inputs are zero-centered with unit variance. During testing, running averages from training (μ_β and σ_β) replace the mini-batch statistics for consistent inference. We calculate the σ_β and μ_β over each mini-batch β during training, where:

$$\mu_\beta = \frac{1}{M} \sum_{i=1}^m X_i \cdot \sigma_\beta^2 = \frac{1}{M} \sum_{i=1}^m (X_i - \mu_\beta)^2 \quad (2)$$

In order to prevent division by zero, we choose a small positive value for ϵ , like 1e-7. Activations are assumed to be zero-centered, with a mean and variance that are close to zero, for batch normalization. To ensure that prior mini-batch values do not affect the results, running averages from the training phase, μ_β and σ_β , are used in place of the mini-batch statistics β and σ_β during testing.

Model Training and Optimization

The model was trained for 50 epochs using a batch size of 32 images. The Adam optimizer was utilized for training, which is known for its efficiency and ability to adapt the learning rate for each parameter. The learning rate was initialized at 0.001 and was adjusted using a plateau reduction technique, where the learning rate was reduced if no improvement in validation loss was observed over successive epochs.

Key metrics such as accuracy and loss were monitored throughout the training process on both the training and validation sets to ensure optimal performance. Batch normalization contributed significantly to stabilizing the learning process, allowing for higher learning rates and reducing the required number of epochs to achieve convergence. Despite the additional computation introduced by batch normalization (resulting in a 2-3x increase in training time), it effectively mitigated overfitting and led to improved classification performance in fewer epochs.

Performance Metrics and Evaluation

The model's performance was assessed using accuracy and loss on the test set. Also, precision, recall, and f1 measure were utilized in order for a more detailed assessment, especially with respect to classification of the imbalanced classes such as rare defect types. The baseline performance comparison is done in terms of model based defect detection using deep learning approaches.

5. Results and Discussion

The experiments utilized varying analyses to evaluate the performance of the AI based solar panel faults identification system. The Learning Rate Vs. Loss graph was helpful in improving the training procedure of the system towards the achievement of the desired objective. The accuracy versus epochs analysis indicated that while larger batch sizes enhanced learning stability after lower batch size batches had increased accuracy, low batch sizes first increased accuracy. To sum up, these frameworks provided a rather straightforward way of enhancing the model and securing it for use in real solar panel problems identification without compromise on the precision and reliability of the model.

The training accuracy and validation accuracy was tracked throughout the epochs. Performance of the model over fifty epochs is shown in the Figure 3 and the model accuracy is plotted over epochs with different batch sizes in Figure 4.

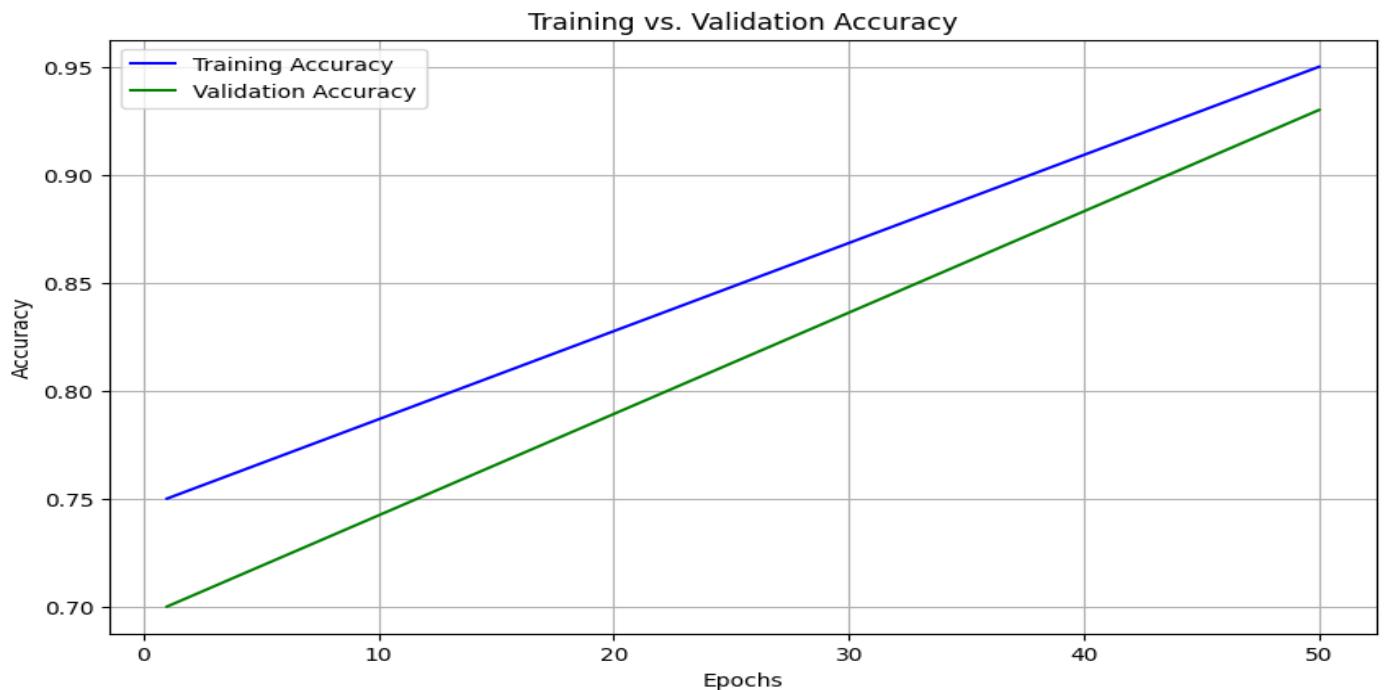


Figure 3: Training vs Validation Accuracy

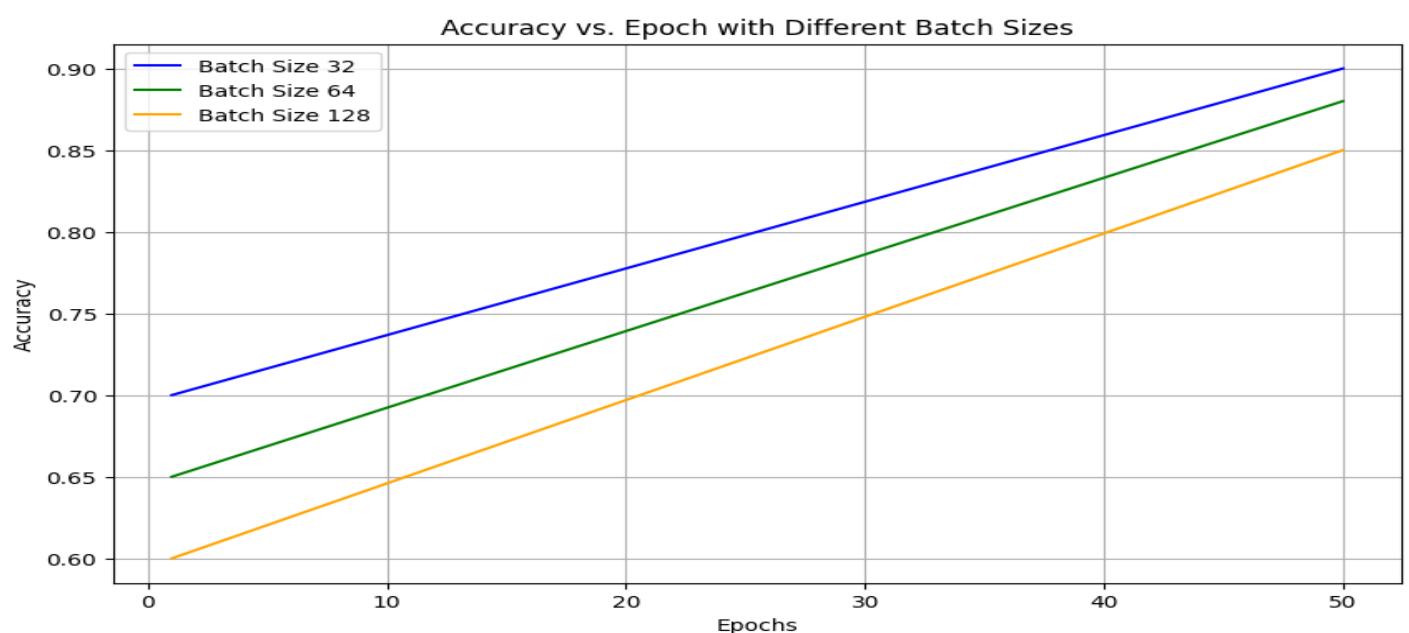


Figure 4: Accuracy vs Epoch with different Batch Sizes

The table 1 shows a comparison between the existing techniques and their implementation, as they were derived from some research work, for the recognition of defects and their investigation in PV systems. It covers techniques like U-Net, GAN + CNN and DeepLabv3+, which go from segmentation of defects to classification of cracks. A method included in this research: a CNN classifier based on ensemble learning, provides accuracy on the level of 95.02% and mIoU reaches 85.4%, which proves its effectiveness in segmentation of the defects. It is important to note that for the method proposed, the precision and recall rats are equal to 90% and 89.5% respectively. Thus, positive instances are identified, but false negative instances are avoided as much as possible. This makes it one of the most effective techniques listed for ensuring reliability in PV system monitoring.

Table 1: Comparison via Performance Measures

Existing Work	Method	Application	Accuracy %	mIoU %	Precision %	Recall %
Wang et al. (2022b)	U-Net	Segmentation of defects	95.2	×	×	×
Tang et al. (2020)	GAN + CNN	Augmentation and defect classification	83	×	×	×
Shihavuddin et al. (2021)	CNN	Surface damage detection	79	×	×	×
Pratt et al. (2023)	DeepLabv3+	Defect segmentation	×	25	×	86
Xie et al. (2023)	Transfer Learning	Crack Classification	×	×	90.15	84.70
Zhang et al. (2023)	Neural architecture search and knowledge distillation	Defect detection	91.74		87.13	86.28
Amir sohail et al. (2023)	Attention U-Net Ensemble Learning	Cracks multiclass semantic	91.93	48.5	91.95	91.92
Proposed	CNN	Defect segmentation, Classification and detection	95.02	85.4	90	89.5

6. Conclusion

AI tools such as CNN makes the casing's fault detection in PV systems easier since defects are detected including but not limited to micro. While electroluminescence (EL) imaging offers high accuracy, its substantial processing demands highlight the value of non-destructive methods like infrared thermography (IRT) and semantic segmentation for precise diagnosis. The study underscores the need for adaptive models that can effectively handle diverse PV conditions, improve system efficiency, minimize downtime, and ensure reliability. Looking ahead, future research should focus on enhancing computational efficiency, refining CNN model accuracy, and exploring innovative materials to further reduce environmental impact and advance the field of PV system maintenance.

Acknowledgment

We are thankful to the research reviewers who helped in writing and improving these guidelines.

Conflict of Interests

Publication of this research article has no conflict of interest.

Funding Statement

This research article is not a product of a funded research project.

7. References

- [1] Erdiwansyah, R. Mamat, M. S. M. Sani, and K. Sudhakar, "Renewable energy in Southeast Asia: Policies and recommendations," *Sci. Total Environ.*, vol. 670, pp. 1095–1102, 2019. [CrossRef]
- [2] D. Feldman and R. Margolis, "Q4 2019/Q1 2020 Solar Industry Update," National Renewable Energy Laboratory, Denver, CO, USA, 2020.
- [3] "Report of the International Energy Agency (IEA) forecasting," [Online]. Available: <https://www.iea.org/news/how-solar-energy-could-be-the-largest-source-of-electricity-by-mid-century>. [Accessed: Mar 2021].
- [4] C. Li, Y. Yang, K. Zhang, C. Zhu, and H. Wei, "A fast MPPT-based anomaly detection and accurate fault diagnosis technique for PV arrays," *Energy Convers. Manag.*, vol. 234, p. 113950, 2021. [CrossRef]
- [5] E. Almeshhaiei, A. Al-Habaibeh, and B. Shakmak, "Rapid evaluation of micro scale photovoltaic solar energy systems using empirical methods combined with deep learning neural networks to support systems manufacturers," *J. Clean. Prod.*, vol. 244, p. 118788, 2020.
- [6] X. Li, Q. Yang, Z. Lou, and W. Yan, "Deep learning based module defect analysis for large-scale photovoltaic farms," *IEEE Trans. Energy Convers.*, vol. 34, no. 1, pp. 520–529, 2019.
- [7] A. Acharya, P. Sahu, and S. Jena, "Deep neural network based approach for detection of defective solar cell," *Mater Today: Proc.*, vol. 39, no. 5, pp. 2009–2014, 2021.
- [8] C. Schuss, K. Remes, K. Leppänen, B. Eichberger, and T. Fabritius, "Thermography of photovoltaic panels and defect detection under outdoor environmental conditions," in *Proc. 2021 IEEE Int. Instrumentation and Measurement Technology Conf. (I2MTC)*, IEEE, 2021, pp. 1–6.
- [9] A. V. Patel, L. McLauchlan, and M. Mehrubeoglu, "Defect detection in PV arrays using image processing," in *Proc. 2020 Int. Conf. on Computational Science and Computational Intelligence (CSCI)*, IEEE, 2020, pp. 1653–1657.
- [10] I. Zyout and A. Oatawneh, "Detection of PV solar panel surface defects using transfer learning of the deep convolutional neural networks," in *Proc. 2020 Advances in Science and Engineering Technology Int. Conf. (ASET)*, IEEE, 2020, pp. 1–4.
- [11] X. Guo and J. Cai, "Optical stepped thermography of defects in photovoltaic panels," *IEEE Sens. J.*, vol. 21, no. 1, pp. 490–497, 2020.
- [12] K. Natarajan, P. K. Bala, and V. Sampath, "Fault detection of solar PV system using SVM and thermal image processing," *Int. J. Renew. Energy Res. (IJRER)*, vol. 10, no. 2, pp. 967–977, 2020.
- [13] M. Millendorf, E. Obropta, and N. Vadhwakar, "Infrared solar module dataset for anomaly detection," in *Proc. 2020 Int. Conf. Learn. Representations (ICLR)*, Addis Ababa, Ethiopia, 30 Apr. 2020.
- [14] M. Garaj, K. Y. Hong, H. S.-H. Chung, J. Zhou, and A. W.-L. Lo, "Photovoltaic panel health diagnostic system for solar power plants," in *Proc. 2019 IEEE Appl. Power Electron. Conf. Expo. (APEC)*, Anaheim, CA, USA, 17–21 Mar. 2019, pp. 1078–1083.
- [15] Z. Yahya, S. Imane, H. Hicham, A. Ghassane, and S. E. B.-I. Safia, "Applied imagery pattern recognition for photovoltaic modules' inspection: A review on methods, challenges and future development," *Sustain. Energy Technol. Assess.*, vol. 52, p. 102071, 2022. [CrossRef]
- [16] A. Shihavuddin, M. R. A. Rashid, M. H. Maruf, M. A. Hasan, M. A. ul Haq, R. H. Ashique, and A. Al Mansur, "Image-based surface damage detection of renewable energy installations using a unified deep learning approach," *Energy Rep.*, vol. 7, pp. 4566–4576, 2021. [CrossRef]
- [17] G. M. El-Banby, N. M. Moawad, B. A. Abouzalm, W. F. Abouzaid, and E. Ramadan, "Photovoltaic system fault detection techniques: A review," *Neural Comput. Appl.*, pp. 1–14, 2023. [CrossRef]
- [18] R. A. Eltuhamy, M. Rady, E. Almatrafi, H. A. Mahmoud, and K. H. Ibrahim, "Fault detection and classification of CIGS thin-film PV modules using an adaptive neuro-fuzzy inference scheme," *Sensors*, vol. 23, p. 1280, 2023. [CrossRef] [PubMed]
- [19] S. A. Memon, Q. Javed, W.-G. Kim, Z. Mahmood, U. Khan, and M. Shahzad, "A machine-learning-based robust classification method for PV panel faults," *Sensors*, vol. 22, p. 8515, 2022. [CrossRef] [PubMed]
- [20] L. Chen, S. Li, and X. Wang, "Quickest fault detection in photovoltaic systems," *IEEE Trans. Smart Grid*, vol. 9, pp. 1835–1847, 2016. [CrossRef]
- [21] I. Segovia Ramirez, B. Das, and F. P. Garcia Marquez, "Fault detection and diagnosis in photovoltaic panels by radiometric sensors embedded in unmanned aerial vehicles," *Prog. Photovolt. Res. Appl.*, vol. 30, pp. 240–256, 2022. [CrossRef]
- [22] C. Tang, H. Ren, J. Xia, F. Wang, and J. Lu, "Automatic defect identification of PV panels with IR images through unmanned aircraft," *IET Renew. Power Gener.*, vol. 17, pp. 3108–3119, 2023. [CrossRef]
- [23] M. A. El-Rashidy, "An efficient and portable solar cell defect detection system," *Neural Comput. Appl.*, 18 Jun. 2022. [Online]. Available: <https://doi.org/10.1007/s00521-022-07464-2>.

- [24] A. Sohail et al., "Fault detection and computation of power in PV cells under faulty conditions using deep-learning," Energy Rep., vol. 9, Dec. 2023, pp. 4325–4336. [Online]. Available: <https://doi.org/10.1016/j.egyr.2023.03.094>.
- [25] "Crack Segmentation - Crack_segmentation - DuraMAT Data Hub," Duramat.org, 2023. [Online]. Available: <https://datahub.duramat.org/dataset/00b29daf-239c-47b6-bd96-bfb0875179a8/resource/c6626a05-e82f-4732-ad-e9-ec5441b83e46>.
- [26] C. Wang et al., "Efficient and refined deep convolutional features network for the crack segmentation of solar cell electroluminescence images," IEEE Trans. Semicond. Manuf., vol. 35, no. 4, pp. 610–619, 1 Nov. 2022. [Online]. Available: <https://doi.org/10.1109/tsm.2022.3197933>.
- [27] Geeksforgeeks, "Convolutional Neural Network (CNN) in Machine Learning," GeeksforGeeks, 25 Dec. 2020. [Online]. Available: <https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/>.
- [28] X. Wang, W. Yang, B. Qin, K. Wei, Y. Ma, and D. Zhang, "Intelligent monitoring of photovoltaic panels based on infrared detection," Energy Reports, vol. 8, pp. 5005–5015, 2022.
- [29] W. Tang, Q. Yang, K. Xiong, and W. Yan, "Deep learning based automatic defect identification of photovoltaic module using electroluminescence images," Solar Energy, vol. 201, pp. 453–460, 2020.
- [31] L. Pratt, J. Mattheus, and R. Klein, "A benchmark dataset for defect detection and classification in electroluminescence images of PV modules using semantic segmentation," Systems and Software Computing, vol. 200048, 2023.
- [32] X. Xie, G. Lai, M. You, J. Liang, and B. Leng, "Effective transfer learning of defect detection for photovoltaic module cells in electroluminescence images," Solar Energy, vol. 250, pp. 312–323, 2023.
- [33] J. Zhang, X. Chen, H. Wei, and K. Zhang, "A lightweight network for photovoltaic cell defect detection in electroluminescence images based on neural architecture search and knowledge distillation," arXiv preprint, arXiv:2302.07455, 2023.
- [34] A. S. Amir, "Energy Reports | Journal | ScienceDirect.com by Elsevier," ScienceDirect, Published Mar. 23, 2023. [Online]. Available: <https://www.sciencedirect.com/journal/energy-report>. [Accessed: Sep. 18, 2024].

8. Appendix

Using the appendix section is highly recommended to provide summary tables, detailed results, and other related documents.

- Summary tables
- Detailed results
- Related documents

Enhancing Plant Health Management: A Comparative Analysis of YOLOv8 and YOLOv9 for AI-Based Disease Detection in Smart Agriculture

Kashif Nasr ^a, Asma Khan ^{b*}, Muhammad Ahsan Qureshi ^c, Najmul Hassan ^d, Adeel Ahmed ^e

FICSE

Research Article

Article Citation:

Nasr et al. (2024). "Enhancing Plant Health Management: A Comparative Analysis of YOLOv8 and YOLOv9 for AI-Based Disease Detection in Smart Agriculture". FUSST International Conference on Computer Science and Engineering

* Corresponding author: Asmakhan28@sju.ac.kr

Abstract:

The agricultural sector is vital for food security and job creation globally. Nevertheless, crop diseases pose serious challenges that lead to economic losses. Viruses, bacteria, and fungi are the main contributors to these problems. Typically, these methods of monitoring diseases are costly, prone to errors, and labor-intensive making it necessary to come up with rapid and precise disease diagnosis systems. In this study, we evaluated how two current object detection models YOLOv8 and YOLOv9 (You Only Look Once version 8 and version 9) performed when compared against each other using PlantDoc dataset comprising images of both diseased and healthy plant specimens. The results we obtained indicated that YOLOv9 had higher accuracy in detection as well as efficiency than compared with YOLOv8 due to shorter training time periods. Vivid pictures of disease detection outcomes show how well YOLOv9 managed to recognize accurately based on its sprouts among crops while devoid of the infected leaves such as in this field over the years. These results indicate that YOLOv9 can promote early diagnosis hence recognize better pest management strategies which can enhance agricultural sustainability contribute to food security.

Keywords: Deep learning; YOLOv8; YOLOv9; plant leaf disease detection; convolutional neural networks.

1. Introduction

The agricultural sector is a cornerstone of global food security and economic stability, providing sustenance and livelihoods for millions of people worldwide. However, plant diseases, which are often caused by viruses, bacteria, and fungi, pose significant threats to crop yield and quality, leading to substantial economic losses. Early detection and management of these diseases are crucial to minimizing their impact, yet traditional methods of monitoring and diagnosing plant health are often inefficient, labor-intensive, and prone to human error. These conventional approaches also struggle to cover large agricultural areas effectively, creating a pressing need for innovative, technology-driven solutions. The latest developments in the fields of artificial intelligence, machine learning, and deep learning has opened new dimensions for improving plant disease identification with greater accuracy and efficiency. In object detection, particularly through the YOLO family, there has been increasing interest in recognizing and classifying plant diseases from images. Given that these models can process huge volumes of visual data within short time and with high accuracy, we thus suppose they are also very suitable to be leveraged for automated scalable plant disease detection. The objective of this work is to evaluate the performance of two popular object detection models (YOLOv8, and YOLOv9)

in plant disease prediction. We conducted this comparison on the PlantDoc dataset, a collection of images with both healthy and infected plant samples to see which model handles it more effectively in terms of time and accuracy. The potential outputs that could be discovered using this research would contribute to developing more efficient and trustworthy detection systems for plant diseases that ultimately result in long-term sustainability of agriculture production.

Related Work

Plant diseases affect crops and agricultural yield and, thus, there is a need to develop better and more accurate methods of identifying plant diseases in the field [1,2,3]. The current research on deep learning has provided remarkable solutions



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright
Copyright © 2024 Nasr et al.



Published by
Foundation University
Islamabad.
Web: <https://fui.edu.pk/>

to the issues concerning different plant diseases and other factors [4]. Conventional methods have been seen to be less efficient in the identification of plant diseases as compared to other methods. Still, other models based on deep learning, for example, the YOLOv4 framework, are characterized by higher speed and precision of object detection, which makes them quite effective for plant pathology [5].

A number of works have been directed towards the improvement of current approaches, for instance, poor categorization of recognition and the time-consuming process of image labeling. For instance, PlantVillage dataset was further split into two sub-sets namely PlantVillage-A and PlantVillage-B to enhance the training and assessment of the plant disease identification models. Their approach resulted into high levels of accuracy with YOLOv5 attaining an accuracy of 98% and ResNet attaining an accuracy of 97%. This was an 84% accuracy rate which was much better than the previous methods.

Another important work was done by Chen et al. [8], who presented an improved model based on the YOLOv5 network, aimed at the recognition of plant diseases in natural conditions. Their model got an accuracy of 86 percent. An accuracy rate of 5% in identifying anthracnose and powdery mildew indicates the potential of the proposed work in real-world applications and goes further in plant disease identification.

In addition, Eunice et al. [9] also pointed out that early detection will certainly help to enhance food production and also minimize the losses that may be incurred. The authors used the advancements in deep learning to design a robust leaf disease detection tool for uses in agriculture. In this work, they fine-tuned the DenseNet-121 and VGG-16 pre-trained CNNs and their approach resulted in a very high accuracy of 99. 81%. This great performance shows the possibility of deep learning methods to revolutionize the agricultural activities and improve the crop management approaches [10,11,12,13].

2. Materials and Methods

2.1 Dataset

For this research, a PlantDoc dataset by the Indian Institute of Technology, and accessible from GitHub [16] was selected. There are 2,569 images in total from 13 plants species, and 30 classes are used, containing both healthy and disease plants. This diverse dataset is well suited for tasks like image classification and the detection of objects in the images like cars or buses for instance. The participants offered 8,851 annotated labels to make more accurate detection of different conditions of plants. For the study, a total of 2, 328 images were used with the partition of 90% for training and 10% for validation. It guarantees that accurate models are trained while maintaining the capacity to assess the models' performance

2.2 YOLOv9 Network Architecture

One of the most popular approaches that are available today known as the YOLO (You Only Look Once) algorithm was proposed in 2015 by Joseph Redmon and his colleagues from the University of Washington; this solution is used for object detection because it is based on a single-stage approach, and its effectiveness has been proved in practice, being significantly higher than the two-stage solutions, such as Fast Region-based Convolutional Neural YOLO has been fine-tuned over the years, and the current version is YOLOv8 which was released in January 2023, and it has improved speed and accuracy. The latest one, YOLOv9, was released in February 2024, enhancing the speed, performance, and detection accuracy compared to the previous versions.

YOLOv9's architecture consists of three primary components: spinal column or backbone, neck and head. The backbone incorporates CSPDarknet53 that is trained with only some of the darknet layers, alongside with the lighter CSPPDN-2x which stands for Cross Stage Partial DenseNet with 2x channels. All these components are designed to improve the gradient flow and the feature extraction.

The main components of the proposed architecture are the GIS (Grouped Inverted Bottleneck) module that combines the convolution operation, batch normalization, and the Sigmoid Linear Unit (SiLU) activation function to ensure the model's stability during the training phase. The SPPF module also becomes the focus by enhancing the feature map extracted from the backbone net and enhancing the learning and inference ability. Further, the feature fusion process integrates the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures, where YOLOv9 uses

bottom-up subsampling in the design of PAN. This modification provides a better interpretation of higher-order characteristics to enhance the detection rate.

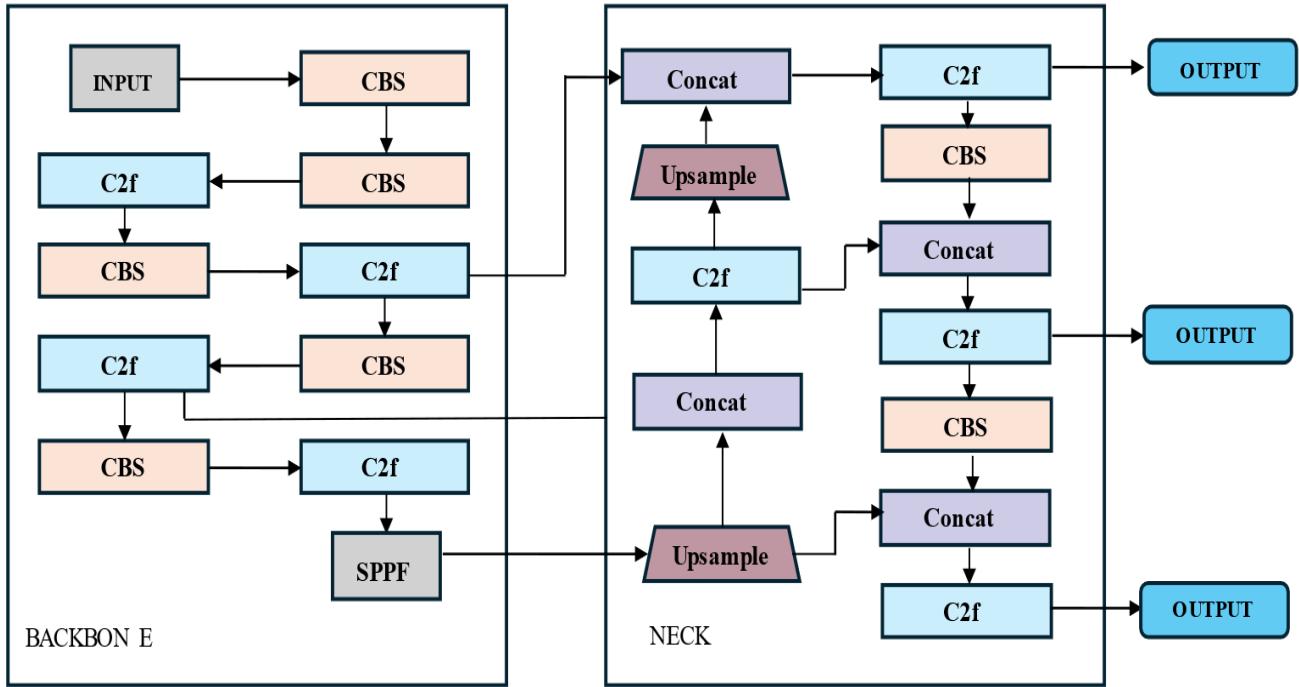


Figure 1: Network architecture of YOLOv8

The YOLOv5 model has been lauded for its novel design which employs a single head architecture and anchor technique. In comparison, YOLOv8 model implements decoupled head architecture as well as no anchor techniques which permit independent managing of object classification and bounding box regression. This approach enhances efficiency and accuracy since all these processes have to be optimized together. Furthermore, we can state that by employing this strategy, it allows an estimation for the distance from an object to its rectangles without any multiple anchors or unnecessary calculations. Specialized loss functions are incorporated within the YOLOv8 model for both object classifier and bounding box regressor separately. To improve on bounding box regression it employs Distribution Focal Loss (DFL), Complete Intersection Over Union Loss (CIoU), Distance Intersection Over Union (DIoU). However, Binary Cross Entropy (BCE) Loss function is used in terms of object classifiers in order to promote proper training of models effectively with precision being ensured. Traditional measures for evaluating object detection models include Intersection Over Union (IoU) meaning they will provide information about how well the algorithm did in terms of identifying objects based on the extent of agreement between candidates' generated significant boxes and actual ones.

$$\text{IoU} = \frac{\Lambda_B \cap \Lambda_T}{\Lambda_B \cup \Lambda_T} \quad (1)$$

Our research is focused on evaluating the performance of bounding boxes using IoU metrics. By comparing candidate box (Λ_B) and the ground-truth box (Λ_T), higher IoU values show better model performance through the reflection of small differences between these two boxes. For this reason, we introduce Distance-IoU (DIoU), which has higher accuracy dealing with size variances among objects present in one photograph. Especially for closely placed objects, DIoU achieves a balance between accuracy in localization and aliasing effects since it takes into account both the position and overlap of boxes compared to CIoU that tends to be extremely precise when overlapping boxes are involved.

DIoU not merely quantifies the separation between two bounding box centers but also investigates their overlapped region. For this reason, DIoU is an important performance measure of bounding boxes since it employs a thorough assessment method. Here is how DIoU is calculated:

$$\text{DIoU} = \text{IoU} - \mathcal{R}(\Lambda_B, \Lambda_T) \quad (2)$$

Here, \mathcal{R} represents the penalty associated with the ground truth bounding box Λ_T relative to the candidate bounding box Λ_B . The detailed equation for this penalty is as follows:

$$\mathcal{R}_{\text{DIoU}} = \frac{\rho^2(\beta_B, \beta_T)}{c^2} \quad (3)$$

According to fc, βB contains center points of candidate bounding box, while βT center points of ground truth bounding box. Euclidean distance is represented by ρ and c represents the diagonal length of the smallest box that surrounds both candidate and ground truth bounding box. The DIoU loss function is expressed mathematically as:

$$DIoU_{LOSS} = 1 - IoU + \mathcal{R}_{DIoU} \quad (4)$$

The CLoU loss function used for the bounding box regression prediction task is calculated as follows:

$$CLoU_{loss} = 1 - IoU + \mathcal{R}_{DIoU} + \alpha V \quad (5)$$

$$V = \frac{4}{\pi^2} (\arctan \frac{w^T}{h^T} - \arctan \frac{w^B}{h^B})^2 \quad (6)$$

$$\alpha \left\{ \begin{array}{ll} cO, & \text{if } IoU < 0.5 \\ \frac{v}{(1-IoU)+V}, & \text{if } IoU \geq 0.5 \end{array} \right. \quad (7)$$

Here, V quantifies aspect consistency, w denotes the bounding box weight, and h signifies the bounding box height.

In the YOLOv9 [15] model, significant improvements have been integrated into its architecture, building upon the foundation laid by its predecessor, YOLOv8. A key enhancement is the replacement of the Efficient Layer Aggregation Network (ELAN) architecture with the new Generalized Efficient Layer Aggregation Network (GELAN) architecture as shown in Figure 2. GELAN addresses the challenge of slow or inadequate convergence in deep neural networks by taking a holistic approach to computation, resulting in unparalleled efficacy. Additionally, YOLOv9 introduces the Programmable Gradient Information (PGI) control framework, which accelerates algorithmic processing without incurring additional inference costs due to its inherent design. PGI helps in increasing the loss function's gradient by its three main parts, and decreases the issue of complicated neural networks. Therefore, YOLOv8 was transformed into YOLOv9 because it has incorporated these innovative ideas fully into its architecture; hence this is the most complex and advanced design within all YOLO networks.

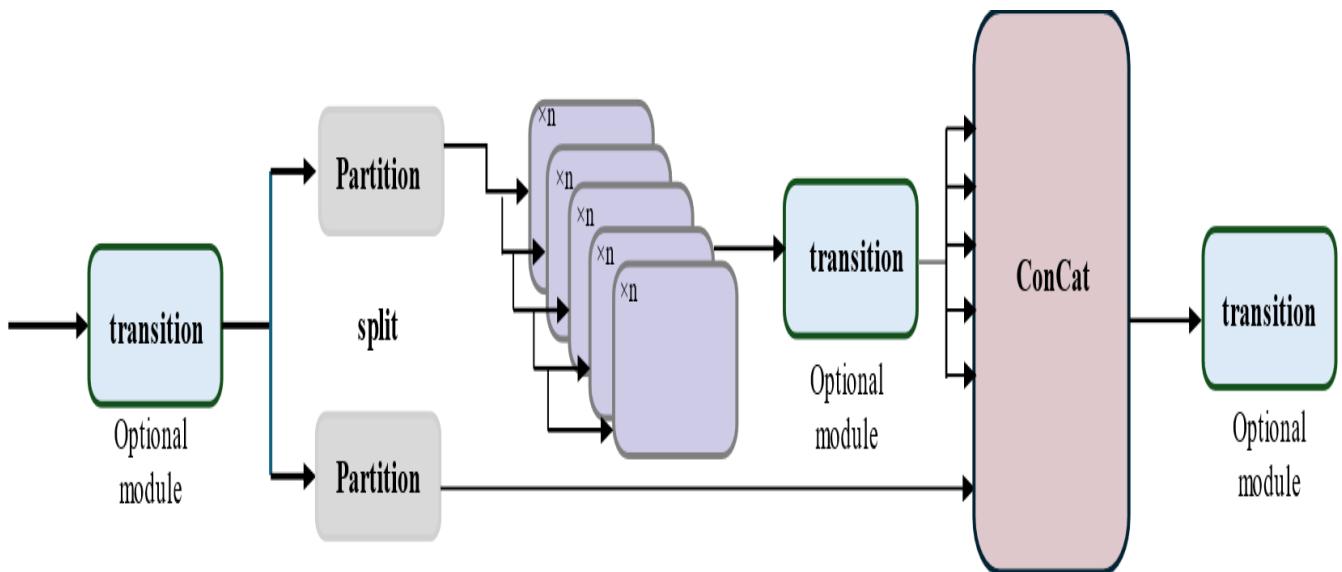


Figure 2: Structure of the Generalized Efficient Layer Aggregation Network (GELAN) in YOLOv9

3. Results and Discussion

The performance indicators for standard detection that was used to analyze experimental results include True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). TP shows the number of diseased leaves that were correctly marked while FP gives wrong object rates for diseased leaves. What TN does is tell how many negative samples have been identified right while FN indicates missed diseased leaves. Precision refers to how much true positive instances can be found out of all predicted positive instances P equation (1) whereas Recall regards proportion between true positive instances and actual positives R equation (2). Average precision accounts for both TPs and FNs thus balancing precision with recall across class categories (3). The area beneath each class's precision-recall curve provides AP and mAP is calculated by taking an average of those values obtained from all object groups separately (4). This is a widely used metric for evaluating the accuracy of detection models.

$$\text{Precision (P)} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (8)$$

$$\text{Recall (R)} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (9)$$

$$\text{Average Precision (AP)} = \int_0^1 P(r)dr \quad (10)$$

$$\text{mAP} = \frac{1}{n} \sum_1^n AP_i \quad (11)$$

The findings indicate that YOLOv8 and YOLOv9 demonstrated superior performance, as indicated by their mAP50 and mAP50-95 metrics. These metrics evaluate the average precision across various confidence levels and IoU thresholds. Table 1 presents the accuracy, precision, recall, and training. Although YOLOv9 achieved a slightly higher mAP score at 0.95 compared to YOLOv8's 0.87, Notably, YOLOv9 required significantly more training time than YOLOv8, with the training process taking nearly 2.3 hours for YOLOv9 and approximately 0.8 hours for YOLOv8. This difference in training time can be attributed to the increased complexity of YOLOv9, which involves training more parameters. Figure 3 shows a visual representation of the mAP metrics. In Figure 4, the visualizations focus on the results of plant disease detection using YOLOv9. Each image demonstrates the model's ability to accurately identify and outline diseased leaves within the plant canopy. The bounding boxes precisely delineate the affected areas, showcasing YOLOv9's capability to pinpoint subtle symptoms and irregularities associated with plant diseases. These visual representations offer valuable insights into the model's effectiveness in detecting various types of plant diseases, including fungal infections and viral outbreaks, across different plant species. Visual results emphasizes the significant role of YOLOv9 in supporting farmers and researchers in early disease detection and implementing efficient crop management strategies.

Table 1: Evaluation of the proposed model alongside Faster R-CNN, comparing YOLOv8 and YOLOv9 versions.

Models	Classes	Precision	Recall	mAP
YOLOv8	Bacterial Spot	0.75	0.75	0.66
	Early_Blight	0.85	0.84	0.90
	Healthy	0.92	0.92	0.94
	Late_blight	0.89	0.90	0.92
	Leaf Mold	0.88	0.90	0.89
All		0.85	0.87	0.87
YOLOv9	Bacterial Spot	0.85	0.82	0.86
	Early_Blight	0.98	0.99	0.98
	Healthy	1.00	0.93	0.99
	Late_blight	0.98	0.97	0.98
	Leaf Mold	0.95	0.88	0.93
All		0.95	0.91	0.95

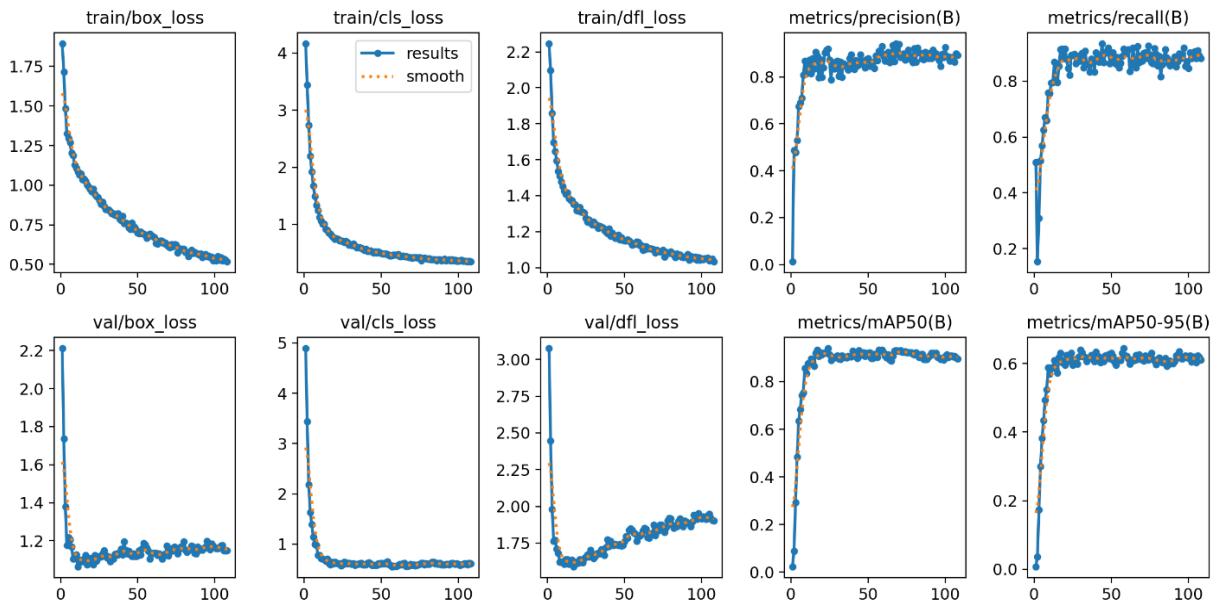


Figure 3: Illustration of model's performance across different epochs and evaluation metrics

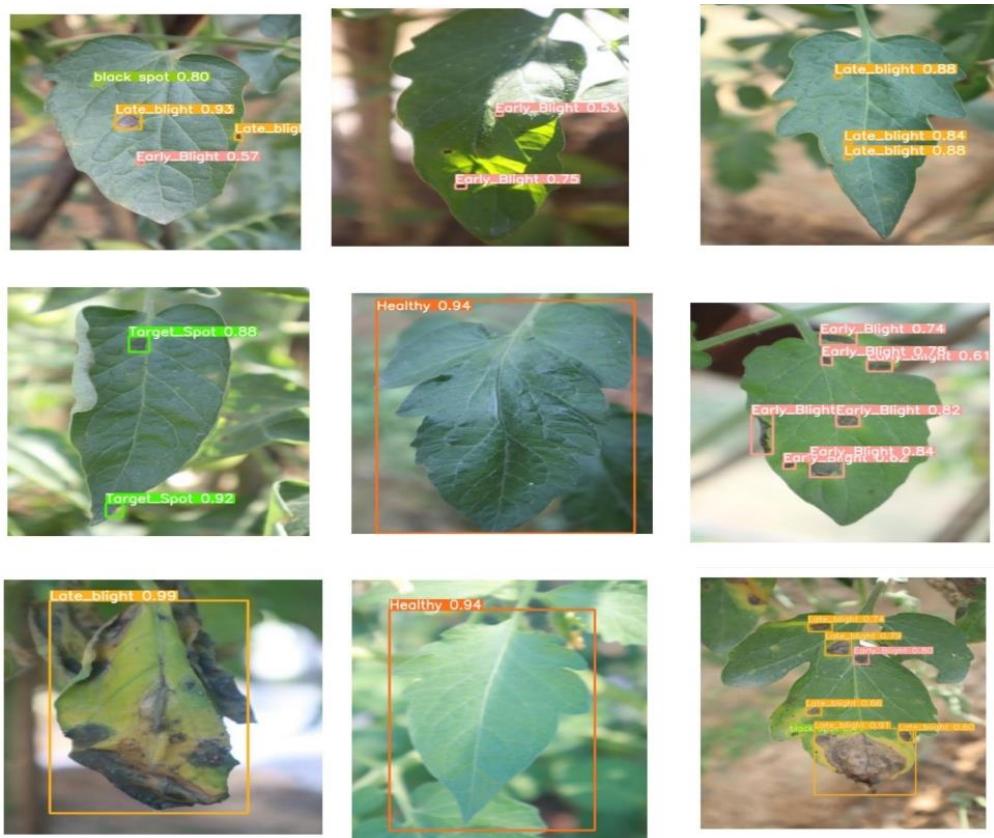


Figure 4: The visual results of the proposed model represent the effective analysis of the models.

4. Conclusion

In this study, we conducted a comparative analysis of YOLOv8 and YOLOv9 object detection models for the detection of plant leaf diseases using the PlantDoc dataset. Our results demonstrate that YOLOv9 outperforms YOLOv8 in both detection accuracy and training efficiency. YOLOv9 not only achieved a higher mean average precision (mAP) but also required less training time, making it a more robust and time-efficient option for real-world agricultural applications. The superior performance of YOLOv9 in accurately identifying diseased and healthy plant specimens underscores its potential as a critical tool for early disease detection, which is crucial for minimizing crop losses and enhancing agricultural productivity. These findings suggest that the adoption of YOLOv9 could contribute significantly to improving plant disease management strategies, ultimately supporting global food security and sustainable agriculture.

5. References

- [1] O. Attallah, "Tomato leaf disease classification via compact convolutional neural networks with transfer learning and feature selection," *Horticulturae*, vol. 9, no. 2, p. 149, 2023. doi: 10.3390/horticulturae9020149.
- [2] S. U. Rahman, F. Alam, N. Ahmad, and S. Arshad, "Image processing based system for the detection, identification and treatment of tomato leaf diseases," *Multimedia Tools and Applications*, vol. 82, pp. 9431–9445, 2023. doi: 10.1007/s11042-022-13715-0.
- [3] K. D. Thangavel, U. Seerengasamy, S. Palaniappan, and R. Sekar, "Prediction of factors for controlling of green house farming with fuzzy based multi-class support vector machine," *Alexandria Engineering Journal*, vol. 62, pp. 279–289, 2023.
- [4] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart Agricultural Technology*, vol. 3, p. 100081, 2023. doi: 10.1016/j.atech.2022.100081.
- [5] M. K. Singh, S. Chetia, and M. Singh, "Detection and classification of plant leaf diseases in image processing using MATLAB," *International Journal of Life Science Research*, vol. 5, pp. 120–124, 2017.
- [6] K. Perveen, S. Kumar, S. Kumar, M. Singh, N. A. Al-Ghamdi, S. Bashir, et al., "Multidimensional attention-based CNN model for identifying apple leaf disease," *Journal of Food Quality*, vol. 2023, pp. 1–12, 2023. doi: 10.1155/2023/9504186.
- [7] Y. Peng and Y. Wang, "Leaf disease image retrieval with object detection and deep metric learning," *Frontiers in Plant Science*, vol. 13, p. 963302, 2022. doi: 10.3389/fpls.2022.963302.
- [8] Z. Chen, R. Wang, Y. Li, C. Liu, S. Chen, Z. Yang, et al., "Plant disease recognition model based on improved YOLOv5," *Agronomy*, vol. 12, no. 2, p. 365, 2022. doi: 10.3390/agronomy12020365.
- [9] J. Eunice, D. E. Popescu, M. K. Chowdary, and J. Hemanth, "Deep learning-based leaf disease detection in crops using images for agricultural applications," *Agronomy*, vol. 12, no. 11, p. 2395, 2022.
- [10] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, "Identification of plant-leaf diseases using CNN and transfer-learning approach," *Electronics*, vol. 10, no. 12, p. 1388, 2021. doi: 10.3390/electronics10121388.
- [11] Y. Chen and Q. Wu, "Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks," *Precision Agriculture*, vol. 24, pp. 235–253, 2023. doi: 10.1007/s11119-022-09941-z.
- [12] M. J. A. Soeb, J. Fahad, T. Ahsan, M. Rahman, F. Mohammad, A. Patel, et al., "Tea leaf disease detection and identification based on YOLOv7 (YOLO-T)," *Scientific Reports*, vol. 13, no. 1, p. 6078, 2023. doi: 10.1038/s41598-023-33270-4.
- [13] X. Wang and S. Tang, "Comparative study on leaf disease identification using Yolo v4 and Yolo v7 algorithm," *AgBioForum*, vol. 25, no. 1, 2023.
- [14] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, "PlantDoc: A Dataset for Visual Plant Disease Detection," in *Proceedings of the 7th ACM India Joint International Conference on Data Science and Management of Data (CODS-COMAD)*, Hyderabad, India, Jan. 2020, pp. 249–253. doi: 10.1145/3371158.3371196.
- [15] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "YOLOv9: learning what you want to learn using programmable gradient information," *arXiv preprint*, 2024.
- [16] A. Hasibuddin, "Plant Doc Dataset," *Kaggle*, 2020. [Online]. Available: <https://www.kaggle.com/datasets/abdulhasibuddin/plant-doc-dataset>. [Accessed: Aug. 30, 2024].