Jatinderkumar R. Saini ·
Shrikant A. Mapari · Amol D. Vibhute ·
Shabana Urooj · Janusz Kacprzyk ·
George Ghinea (Eds.)

Communications in Computer and Information Science 2538

Advancements in Machine Learning

First International Conference, ICCAML 2024
Pune, India, February 28–29, 2024
and Second International Conference, ICCAML 2025
Pune, India, February 25-26, 2025, Proceedings

Springer

Communications in Computer and Information Science 2538

Rationale
The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics
The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors
Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (http://link.springer.com) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as post-proceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at http://link.springer.com/bookseries/7899. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process
The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing
CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start
To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com

Jatinderkumar R. Saini · Shrikant A. Mapari ·
Amol D. Vibhute · Shabana Urooj ·
Janusz Kacprzyk · George Ghinea
Editors

Advancements in Machine Learning

First International Conference, ICCAML 2024
Pune, India, February 28–29, 2024
and Second International Conference, ICCAML 2025
Pune, India, February 25–26, 2025, Proceedings

Springer

Editors
Jatinderkumar R. Saini
Symbiosis Institute of Computer Studies and Research
Pune, Maharashtra, India

Shrikant A. Mapari
Symbiosis Institute of Computer Studies and Research
Pune, Maharashtra, India

Amol D. Vibhute
Symbiosis Institute of Computer Studies and Research
Pune, Maharashtra, India

Shabana Urooj
Princess Nourah Bint Abdulrahman University
Riyadh, Saudi Arabia

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland

George Ghinea
Brunel University
London, UK

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-98137-1 ISBN 978-3-031-98138-8 (eBook)
https://doi.org/10.1007/978-3-031-98138-8

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

Preface

It is a matter of great privilege to have been tasked with the writing of this preface for the proceedings of The First and Second International Conferences on Current Advancements in Machine Learning (ICCAML 2024 & 2025). The conference aimed to provide an excellent international forum for emerging and accomplished research scholars, academicians, students, and professionals in the areas of computer science and engineering to present their research, knowledge, new ideas, and innovations. The conference was held February 28–29, 2024 and February 25–26, 2025 at Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India.

There were three tracks i) Modernization in Machine Learning, ii) Machine Learning for Data Analytics, and iii) Machine Learning for automation. Research submissions in these three areas were received. The Program Committee of ICCAML 2024 & 2025 is extremely grateful to the authors from different countries that showed an overwhelming response to the call for papers, submitting 173 papers. The entire review team (Technical Program Committee members) expended tremendous effort to ensure fairness and consistency during the selection process, resulting in the best-quality papers being selected for presentation and publication. It was ensured that every paper received at least three, and in most cases four, single-blind reviews. Checking of similarities and AI writing detection was also done based on international norms and standards. After a rigorous peer review, 19 papers were accepted, with an acceptance ratio of 10.55%. The papers are organized according to the tracks of conference.

The proceedings of the conference are published as one volume in the Communications in Computer and Information Science (CCIS) series by Springer, and are also indexed by ISI Proceedings, DBLP, Ulrich's, EI-Compendex, SCOPUS, Zentralblatt Math, MetaPress, and Springerlink. We, in our capacity as volume editors, convey our sincere gratitude to Springer for providing the opportunity to publish the proceedings of ICCAML 2024 & 2025 in their CCIS series.

ICCAML 2024 & 2025 provided an excellent international virtual forum for the conference delegates and a platform for scientists, researchers, engineers, and students to exchange their ideas for betterment of the community and society in the areas of advanced technologies in Machine Learning.

The ICCAML 2024 & 2025 conferences received research papers from various domains with a Machine Learning background. Data is processed by using Machine Learning techniques; the outcome of the model and its need is perpetual. The perpetual system keeps track of the need for information continuously and updates are made automatically. Here the basic challenge for researchers, scientists, academicians, developers, and students is how to provide the information as and when needed or periodically or perpetually to fulfill the requirements of upcoming systems, legacy systems, operational databases, external sources, etc. for real-time problem solving and decision making.

The editors are indebted to all the members of the organizing committees, the authors, and the reviewers for their support in making this conference successful. Special thanks

to the staff of Springer for their help and cooperation. Last but not least, the editors profusely thank all who directly or indirectly helped us in making ICCAML 2024 & 2025 a grand success and allowed the conference to achieve its goal, academic or otherwise.

May 2025

Jatinderkumar R. Saini
Shrikant A. Mapari
Amol D. Vibhute
Shabana Urooj
Janusz Kacprzyk
George Ghinea

Organization

Chief Patron

S. B. Mujumdar	Symbiosis International (Deemed University), India

Patrons

Vidya Yeravdekar	Symbiosis International (Deemed University), India
Ramakrishnan Raman	Symbiosis International (Deemed University), India

General Chair

Jatinderkumar R. Saini	Symbiosis International (Deemed University), India

General Co-chairs

Shrikant A. Mapari	Symbiosis International (Deemed University), India
Shabana Urooj	Princess Nourah Bint Abdulrahman University, Saudi Arabia

Program Chair

Sarika Sharma	Symbiosis International (Deemed University), India

Technical Program Committee Chairs

Amol Vibhute	Symbiosis International (Deemed University), India
Sandeep Gaikwad	Symbiosis International (Deemed University), India

Sponsorship Committee Chairs

Parag Ravikant Kaveri	Symbiosis International (Deemed University), India
Madhu Arora	Symbiosis International (Deemed University), India
Sanket Kurdukar	Symbiosis International (Deemed University), India

Publicity Committee Chairs

Rajesh Kumar Dhanaraj	Symbiosis International (Deemed University), India
Prafulla Bafna	Symbiosis International (Deemed University), India
Aniket Nagane	Symbiosis International (Deemed University), India

Finance Chairs

Parag Ravikant Kaveri	Symbiosis International (Deemed University), India
Deepali Suryavanshi	Symbiosis International (Deemed University), India

Contents

Learning with Limited Data: A Machine Learning Approach for Heart Disease Prediction

V. N. Manjunath Aradhya(✉), P. Shiva Prasad, and C. S. Chaithra

Department of Computer Applications, JSS Science and Technology, University, Mysuru 570006, India
aradhya@jssstuniv.in

Abstract. Cardiovascular diseases (CVDs) continue to be a significant global health challenge, underscoring the need for accurate risk prediction models to improve early detection and prevention. In this work two datasets from different sources are combined; the necessary features from the two sources were used to predict instances of heart disease depending on cholesterol, blood pressure, fasting blood sugar, and many other factors. The validation set comprised 1,918 instances and were trained to evaluate a number of machine learning algorithms such as Logistic Regression, Naive Bayes, Elastic Net (E-Net), Bagging ensemble model, Sparse SVM and Gaussian process regression. The main contribution of the work is to understand how the machine learning models learn when the data is very limited, say two samples (one sample from each class). To check the robustness, the analysis was carried out on two-samples training sets from 500 distinct combinations of heart diseases and normal cases.

Keywords: Limited Data · Heart Disease Prediction · Machine Learning · Good Health

1 Introduction

CVDs are still the leading global killers accounting for at least 17% of all human fatalities. About 9 million people die each year, this burden is worse felt in low and middle-income countries where health care access is relatively limited [1]. Estimations show that CVD would increase in Asia by more than 91% by 2050 in terms of mortality rates with ischemic heart disease (IHD) and stroke being the main causes [2]. According to Global Burden of Disease (GBD) 2016 analysis of heart disease trends highlights the growing impact of risk factors that can be controlled, such as high blood pressure, poor diet, and pollution. These risk factors are leading to more deaths and contributing to the rise in disability-adjusted life years (DALYs), years lived with disability (YLD), and years of life lost (YLL) due to preventable early deaths [3].

ECG, stress test and other related clinical diagnostic tools used in clinical diagnosis of CVD are known but have many drawbacks especially in early stages of CVD when the patients are asymptomatic. This has in turn created a need to apply artificial intelligence (AI) and machine learning (ML) in the health care sector [4]. In healthcare adopting the

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 1–12, 2026.
https://doi.org/10.1007/978-3-031-98138-8_1

use of ML models as a subfield of AI, predictions are made using clinical data including the age of the patient, cholesterol levels, blood pressure, levels of glucose among other things [5]. A report on the role of artificial intelligence in healthcare revealed that AI can diagnose, and predict diseases, improve and even automate the clinical decisions [6].

In this paper, we use different ML methods to predict heart disease using two dataset which are publicly available dataset, we combined the dataset and the final dataset obtained of 1,918 records with total 12 attributes. Basic clinical descriptors are also provided in the set including age, cholesterol and blood pressure measures, components which are vital in determining heart disease risks. Several training model algorithms, namely Logistic Regression, Naive Bayes, Elastic Net, Sparse SVM, Bagged Ensemble method and Gaussian Process Regression are tested out with two different training approach, where the first approach was with limited training data [25, 26]. This approach assists in evaluating how those models can perform in a clinical environment where there sometimes is scarce data available. For this purpose, we incorporate 500 different distinct two-sample training sets in an experiment to test the robustness and precision of the models. Thus, the study aims to find the best ways to apply machine learning methods on the data identified of cardiovascular disease risk assessment and treatment strategies.

2 Related Work

This section presents the work focused on clinical decision support systems that use machine learning and deep learning in disease diagnosis particularly heart diseases. In [7], the authors present a robust ensemble machine intelligence framework for CVD classification. The proposed model uses combination algorithms like Naive Bayes, Random Forest, SVM, and XGBoost, this model reached 96.75% on the Mendeley dataset, 93.39% on IEEE DataPort and 88.24% on the Cleveland dataset respectively. In [8] Omkari and Shaik applied Two-Layered Voting (TLV) Framework for coronary artery disease prediction by analyzing big data from Kaggle and UCI. The soft and hard voting strategies are applied with classifiers such as Decision Tree and Support Vector Classifier tuned using GridSearchCV. The model achieved 99.03% accuracy on the UCI dataset and 88.09% on Kaggle dataset. Hammoud et al. [9] performed a comparative study of machine learning algorithms for coronary heart disease prediction, where they compared multiple algorithms and highlighted their strengths in terms of prediction accuracy. Mienye and Jere [10] proposed an optimized ensemble learning approach for heart disease prediction, combining Bayesian optimization for hyperparameter tuning and Shapley additive explanations (SHAP) for interpretability. This approach combined the AdaBoost, random forest, and XGBoost models and achieved high specificity and sensitivity on both the Cleveland and Framingham datasets, with the XGBoost model performing the best. Bhowmik et al. [11] utilized the Cleveland dataset from the UCI Machine Learning Repository, which contains 70,000 patient records with 12 features, to predict heart disease using machine learning algorithms, including Logistic Regression, Random Forest, and Support Vector Machines. The models were evaluated on precision, accuracy, recall, F1-score, and ROC-AUC. Logistic Regression outperformed the others, achieving the highest ROC-AUC score and demonstrating a balanced

trade-off between true positives and false positives. In [12], El-Sofany et al. introduced a machine learning framework for the prediction of heart disease, employing feature selection techniques like chi-square, ANOVA, and mutual information for the optimization of classification accuracy. The class imbalance issue was addressed with SMOTE and models such as XGBoost and Random Forest were used; a peak accuracy of 97.57% was achieved. Explanatory AI tools like SHAP were used to enhance the transparency of models and help determine significant predictors in heart disease diagnosis. Mohan et al. [13] developed an efficient mode to predict the odds of heart disease with increased accuracy using Hybrid Random Forest with Linear Model (HRFLM). Applying the different feature combinations and classification approaches, the above model obtained an accuracy of 88. 7%. Dubey A. K. et al. [14] used different ML algorithms for classification of heart disease. Cleveland and Statlog datasets from the UCI Machine Learning repository were used to build and evaluate the model. The outcome of the experiment reveals that both the classifier models, namely LR and SVM have higher levels of performance with 89% accuracy in the Cleveland dataset and 93% accuracy in the Statlog dataset. In [15] Maini et al. proposed a machine learning model that could assist CVD early diagnosis that was conducted in South India. Authors used five different algorithms of ML by training them on a dataset comprising of 1670 medical records. As a result, the best one was identified as the Random Forest method providing 93.8% accuracy of classifications, 92.8%. In [16] Gupta et al. used different supervised machine learning models to predict Cardiac disease and the data set used was from UCI. They include comparing logistic regression with K-nearest neighbor, support vector machines, decision tree and naïve Bayesian. Logistic Regression produces the highest accuracy, precision, and recall for cardiac disease prognosis.

3 Proposed Methodology

In this study, several machine learning algorithms were employed to develop models for predicting heart disease. We chose these algorithms because they are well-suited for clinical data and offer reliable results, even with small datasets. In addition to these algorithms, various training methods were utilized to evaluate how well these models perform. This approach of using multiple algorithms along with appropriate training techniques enables us to assess their effectiveness in predicting heart disease. The following sections give a brief description of each algorithm employed in the study.

3.1 Logistic Regression

Logistic Regression is a type of linear model which is used in binary classification useful for predicting the presence or absence of conditions like heart disease, based on independent variables. The ability of Logistic Regression to model the relationship between the features and the target variable using a linear approach makes it both interpretable and straightforward. The logistic function ensures that the predictions fall in a bounded range of 0 and 1, which is suitable for estimating probabilities in health-related predictions[15].

3.2 Naive Bayes Regression

Naive Bayes is a probabilistic classifier based on Bayes' Theorem, which simplifies computations by assuming that the features are conditionally independent. This makes it efficient for a wide range of classification problems. In general applications, Naive Bayes calculates the posterior probability of an outcome based on each input feature, such as demographic, clinical, or textual data, allowing for quick and scalable predictions. Its efficiency and the ability to deal with categorical and continuous data make it exceptionally well-suited to handle healthcare-related datasets where fast predictions are vital, and complex relationships between features remain conditionally independent [16].

3.3 Elastic Net Regression

Elastic Net combines the strengths of Lasso (L1) for feature selection and Ridge (L2) for managing multicollinearity, making it ideal for datasets with many interrelated features, such as medical data. It selects the most significant predictors and helps prevent overfitting, leading to more reliable models. With so many correlated features in a medical dataset, the Elastic Net method is quite well-suited for a balanced approach: it can retain the most important predictors while downplaying the influence of less relevant ones. This increases the model's robustness and the likelihood that it generalizes well when applied to new, unseen data [17, 21]. The objective function is:

$$\min_{\beta}\left(\frac{1}{2N}\|y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_{2\|}\beta\|_2^2\right) \tag{1}$$

where λ1 and λ2 control the L1 and L2 penalties.

3.4 Bagged Ensemble Model (Bagging)

The Bagged Ensemble Model, commonly known as Bagging, is an ensemble learning technique that helps to maintain the stability of machine learning models by minimizing their variance. It means creating several copies of the bootstrapped datasets from the initial data then training different models with these data copies and finally generating the final prediction based on the result of the models. This model especially helpful in medical datasets, where data variability can lead to overfitting in individual models. Bagging reduces this risk by averaging the predictions of multiple models, thus providing a more reliable and robust output that can better handle the inherent noise and variability often present in health-related data [18, 22]. The update rule is as follows:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \tag{2}$$

where B represents the number of bootstrap samples, and $f_b(x)$ is the prediction from the b^{th} model.

3.5 Sparse Support Vector Machine (Sparse SVM)

Sparse Support Vector Machine (Sparse SVM) is an adaptation of traditional SVM that enhances simplicity and sparsity by incorporating L1 regularization. This approach reduces the number of support vectors required for classification, allowing the model to identify key predictors while disregarding irrelevant features. By focusing on the most significant features, Sparse SVM helps in building a more efficient and interpretable model, reducing complexity while still maintaining predictive accuracy. This is especially important when trying to identify crucial health indicators from potentially noisy and high-dimensional clinical data [19, 23]. It is represented as follows:

$$\min_{w,b}\left(\frac{1}{2}\|w\|^2 + c\sum_{i=1}^{n}\max(0,1 - y_i(w^T x_i + b))\right) + \lambda\|w\|_1 \tag{3}$$

Where,

- $\frac{1}{2}\|w\|^2$: Like in standard SVM, this term maximizes the margin between the classes by keeping the weights small.
- $c\sum_{i=1}^{n} max(0,1 - y_i(w^T x_i + b)$: This penalizes misclassifications, similar to regular SVM.
- $\lambda\|w\|_1$: This is the new term that forces some weights to zero, leading to a model that uses fewer features, which simplifies the model and makes it more interpretable.

3.6 Gaussian Process Regression (GPR)

GPR is a non-parametric probabilistic model which gives predictions with a measure of uncertainty. GPR works in the prediction of heart diseases by using a kernel function to model the relation between features like blood pressure, age, and cholesterol, hence having the capability to capture complicated patterns within the data. Its ability to provide both predictions and an uncertainty measure aligns with the need for models that not only predict outcomes but also indicate the confidence level in predictions, which is crucial in sensitive domains like healthcare. The probabilistic nature of the model offers a range of possible outcomes for the certainty of each prediction [20, 24].

4 Experiment Results and Discussion

4.1 Dataset Description

The current study analyzes two publicly available data sets to comprehensively predict heart disease. The sample data set used in the current study is the heart illness dataset drawn from one of the multi-specialty hospital in India [13] and consist of 1000 subjects and 12 features related to cardiovascular disease. The second dataset is named heart disease datasets from UCI Machine Learning Repository [14], which is integrated with the Cleveland, Hungarian, Switzerland, Long Beach VA, and Stalog (Heart) datasets. The integration is based on 11 common features of overall dataset. The dataset includes 1190 instances. Considering both the dataset, we combined the dataset which was collected and then combined dataset had 2190 instances, then the 272 duplicate instances was

removed. The purpose of combining such data sets was to create a reliable and thorough resource for the prediction of the heart disease. The updated dataset contains 1989 records with 12 attributes it provides the strength of the two sources. These attributes include basic demographic details like age and gender, along with key clinical factors such as resting blood pressure, cholesterol levels, chest pain type, and exercise-induced angina. This enriched dataset allows better analysis and improves the accuracy of the classification of machine learning models.

4.2 Results

In the first approach, the dataset was divided into training and testing samples, with the minimum number of training samples of two and up to 100 samples. The aim was to check how the models work when the amount of training data increases gradually. Since medical datasets may have different sizes, it is important to know how the models perform on a dataset of a certain size. It is useful for constructing the efficient heart diseases predication models because it is likely to have fewer data to arrive at high accurate diagnostic results. This approach was intended to monitor how accuracy change with the increase in the data size. Table 1 shows the accuracy for varying samples. Based on the findings, it is noted that models such as Navie Bayes, Elastic Net Regression, Bagged Ensemble model and Sparse Support Vector Machine had the highest level of accuracy between 85.5% to 87.4% was achieved at 100 samples trained, where as other models such as Logistic Regression and Gaussian Process Regression model achieve accuracy of 86.18 and 70.23% at 80 and 50 samples trained.

In the second approach, the models were trained with only two data samples and generated 500 such unique combinations and wanted to see how models perform when training such limited samples. Some of them can potentially preserve these patterns which results in higher accuracy and ability to distinguish between classes (e. g., healthy vs. diseased) based on the features present in the dataset. This approach mimics practical application situations in which data is limited by evaluating the capacity of the model to generate accurate predictions with limited data inputs. Figures 1, 2, 3, 4, 5 and 6, shows the accuracy distribution of these combinations has been presented. From the results it is clear that bagged ensemble model outperforms all other models with an impressive 86.69% of accuracy, when two samples trained. To evaluate the model performance more systematically, the maximum, minimum, and average accuracy is also calculated for each model, which is shown in Table 2.

These findings are consistent with the work of other researchers in the field. For example, [7] showed that ensemble methods, like Random Forest and XGBoost, offer high accuracy in classification tasks with larger datasets, while our study confirms that Bagged Ensemble is robust even with minimal data. In addition, [9] shows the application of Hybrid Random Forest models for better accuracy, which can be seen to support the strength of ensemble models in heart disease prediction, particularly when data is limited. However, it is noteworthy that models like Logistic Regression and Gaussian Process Regression performed less well with fewer training samples, consistent with results in studies such as [12], where Logistic Regression was also less effective in low-data scenarios. In conclusion, the results of this study show that Bagged Ensemble and Sparse Support Vector Machine are very efficient in heart disease prediction

Table 1. Accuracy for varying samples

Training samples	Logisitic Regression	Navie Bayes Regression	Elastic Net Regression	Bagged Ensemble Model	Sparse SVM	Gaussian Process Regression
2	56.68	56.68	56.68	56.68	56.68	56.68
3	49.34	56.71	55.61	61.82	63.60	56.71
4	45.55	62.59	50.78	56.32	60.91	60.50
5	69.36	56.50	67.06	61.84	77.99	60.63
10	48.74	70.44	67.13	60.63	80.87	65.51
20	72.81	77.76	67.07	84.98	83.29	56.74
30	85.27	84.69	82.99	84.69	83.79	59.95
40	74.60	84.61	81.36	86.31	76.51	56.70
50	83.83	81.58	85.97	82.28	84.63	**70.23**
60	82.18	81.86	85.79	86.11	83.69	67.70
70	79.16	83.27	83.92	84.41	80.19	69.37
80	**86.18**	85.09	86.01	86.72	86.39	67.68
90	82.76	84.08	82.38	85.12	81.67	67.56
100	85.64	**85.53**	**87.29**	**87.40**	**86.57**	59.13

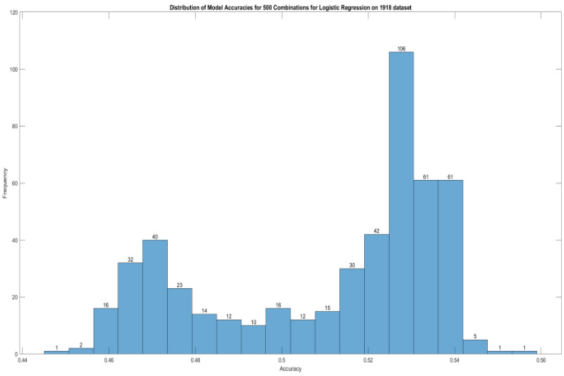


Fig. 1. Distribution of Logistic Regression

models, especially when the data is limited. These models perform well as the size of the dataset increases and retain the capability to generate reliable predictions with smaller datasets. This insight is very important for developing heart disease prediction systems that can function efficiently even in settings where data availability is restricted.

The study findings also demonstrate the usefulness of machine learning models, including Bagged Ensemble and Sparse Support Vector Machine, in practical clinical environments with limited data. Even with small datasets, these models retain

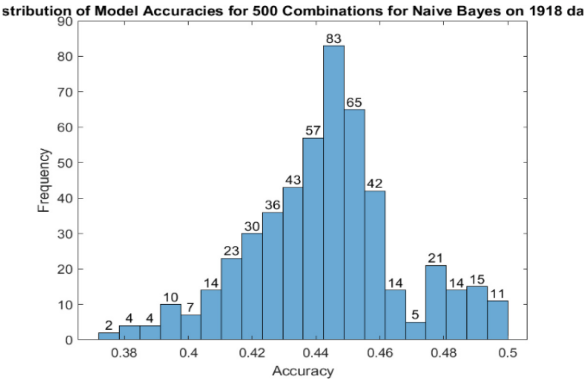


Fig. 2. Distribution of Navie Bayes

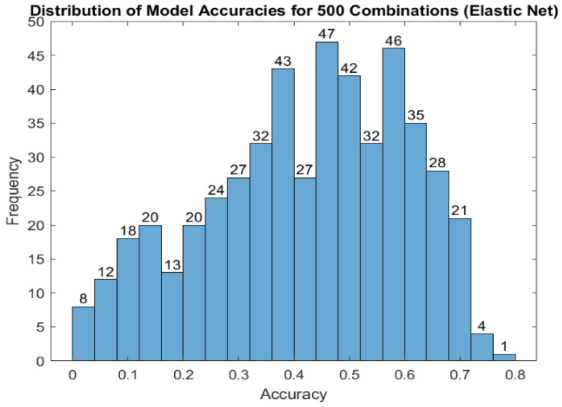


Fig. 3. Distribution of Elastic Net

the extremely high predictability that would be valuable for healthcare professionals deployed in poor resource settings such as a rural clinic or emergency care unit. These models can help clinicians make informed decisions and avoid severe heart disease outcomes by enabling early diagnosis and intervention with minimal input data. This study emphasizes the importance of predictive models in overcoming data scarcity and improving diagnostic capabilities in real-world healthcare settings. Ultimately, integrating these models into clinical practice can enhance early disease detection and intervention, improving patient outcomes even in low-resource environments.

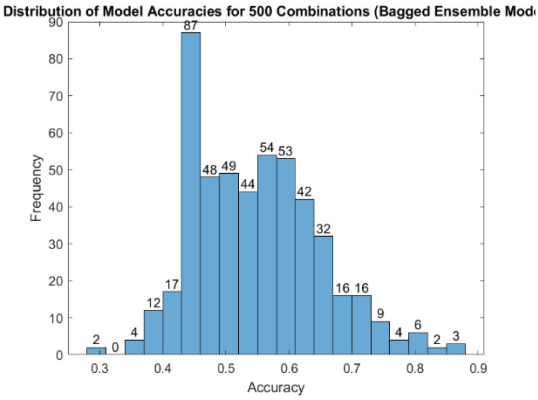


Fig. 4. Distribution of Bagged Ensemble

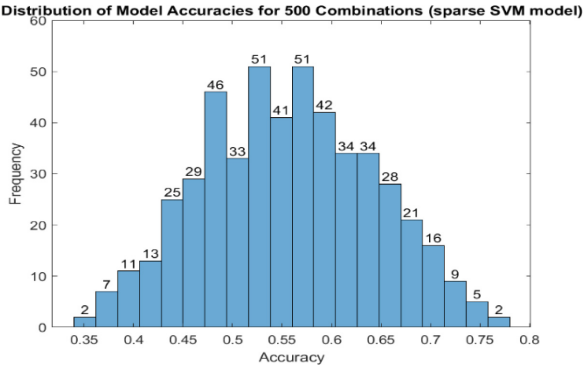


Fig. 5. Distribution of Spared SVM

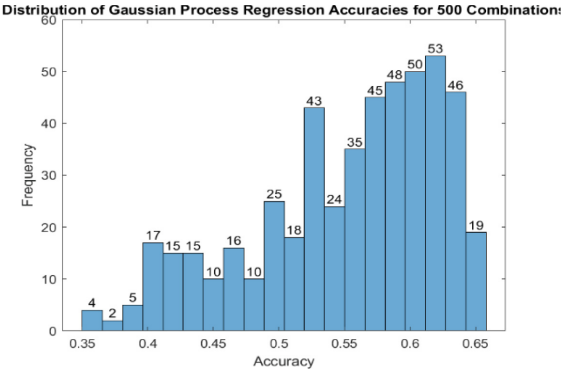


Fig. 6. Distribution of Gaussian Process Regression

Table 2. Accuracy - Average, Maximum and Minimum

Models	Average	Maximum	Minimum
Logistic Regression	51.04	55.90	**44.57**
Navie Bayes Regression	44.32	49.84	37.63
Elastic Net Regression	42.05	78.18	1.57
Sparse SVM	**55.62**	77.87	35.70
Bagged Ensemble Model	54.73	**86.69**	29.28
Gaussian Process Regression	55.22	65.66	35.44

5 Conclusion

This work demonstrates the enrichment of both large and small training datasets for better predictive capabilities over models designed for heart diseases. From the results of both training approach, we can conclude that even with limited (say two), strategically selected combinations may lead to a profound predictive accuracy. These results indicate that very small training data (say two) is much sufficient to predict the early signs of diseases. Future work might be the improvement of binary models using more discrete attributes by enhancing early predictions and intervention on time to avoid more serious health events such as sudden attacks.

Although the findings of this study are promising, there are a number of limitations that need to be addressed. The dataset used here does not fully represent the diversity of patient populations encountered in real-world clinical settings. In future research, the dataset needs to be expanded to include more diverse demographics and health conditions for better generalization of the model. Additionally, advanced feature engineering and data preprocessing methods could further improve the accuracy of the models.

Further studies should also look into hybrid models and real-time data from wearable devices, which may further improve the accuracy of predictions and allow for timely intervention. Longitudinal studies are needed to assess the long-term performance of these models in tracking the progression of heart disease. Finally, improving the interpretability of these models will be important for their integration into clinical practice, ensuring they are useful for healthcare practitioners.

References

1. World Health Organization. Cardiovascular diseases (CVDs) (2021). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. Goh, R.S.J., Tan, C.J., Chong, K.L., et al.: The burden of cardiovascular disease in Asia from 2025 to 2050: a forecast analysis for East Asia, South Asia, South-East Asia, Central Asia, and high-income Asia Pacific regions. Lancet Region. Health – Western Pacific **49**, 101138. (2024). https://doi.org/10.1016/j.lanwpc.2024.101138
3. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE)

for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet **390**(10100), pp. 1260–1344 (2017). https://doi.org/10.1016/S0140-6736(17)32130-X
4. Hussain, K., Tariq, A., Gill, A.Y.: Role of artificial intelligence in cardiovascular health care. J. World Sci. **2**(4), 583–591 (2023). https://doi.org/10.58344/jws.v2i4.284
5. Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. N. Engl. J. Med. **380**(14), 1347–1358 (2019). https://doi.org/10.1056/NEJMra1814259
6. Topol, E.J.: High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. **25**(1), 44–56 (2019). https://doi.org/10.1038/s41591-018-0300-7
7. Doppala, B.P., Bhattacharyya, D., Janarthanan, M., Baik, N.: A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques. J. Healthcare Eng. **2022**, 2585235 (2022). https://doi.org/10.1155/2022/2585235
8. Omkari, D.Y., Shaik, K.: An integrated two-layered voting (TLV) framework for coronary artery disease prediction using machine learning classifiers. IEEE Access **12**, 56275–56290 (2024). https://doi.org/10.1109/access.2024.3389707
9. Hammoud, A., et al.: Coronary heart disease prediction: a comparative study of machine learning algorithms. J. Adv. Inform. Technol. (2024)
10. Mienye, I.D., Jere, N.: Optimized ensemble learning approach with explainable AI for improved heart disease prediction. Information **15**(7), 394 (2024). https://doi.org/10.3390/info15070394
11. Bhowmik, P.K., et al.: Advancing heart disease prediction through machine learning: techniques and insights for improved cardiovascular health. British J. Nurs. Stud. **4**(2), 35–50 (2024). https://doi.org/10.32996/bjns.2024.4.2.5
12. El-Sofany, H.F.: Predicting heart diseases using machine learning and different data classification techniques. IEEE Access **12**, 106146–106160 (2024). https://doi.org/10.1109/ACCESS.2024.3437181
13. Mohan, S., Thirumalai, C., Srivastava, G.: Effective heart disease prediction using hybrid machine learning techniques. IEEE Access **7**, 81542–81554 (2019). https://doi.org/10.1109/ACCESS.2019.2923707
14. Dubey, S.K., Sinha, S., Jain, A.: Heart disease prediction classification using machine learning. Int. J. Invent. Eng. Sci. **10**(11), 1–6 (2023). https://doi.org/10.35940/ijies.b4321.11101123
15. Maini, E., Venkateswarlu, B., Maini, B., Marwaha, D.: Machine learning-based heart disease prediction system for Indian population: an exploratory study done in South India. Med. J. Armed Forces India **77**(1), 45–51 (2021). https://doi.org/10.1016/j.mjafi.2020.10.013
16. Gupta, C., Saha, A., Reddy, N.V.S., Acharya, D.: Cardiac disease prediction using supervised machine learning techniques. J. Phys: Conf. Ser. **2161**, 012013 (2022). https://doi.org/10.1088/1742-6596/2161/1/012013
17. Doppala, B.P., Bhattacharyya, D.: Cardiovascular_Disease_Dataset. Mendeley Data **V1** (2021). https://doi.org/10.17632/dzz48mvjht.1
18. Fedesoriano. Heart failure prediction dataset (2021). https://www.kaggle.com/fedesoriano/heart-failure-prediction
19. Anshori, M., Haris, M.S.: Predicting heart disease using logistic regression. Knowl. Eng. Data Sci. **5**(2), 188–196 (2022). https://doi.org/10.17977/um018v5i22022p188-196
20. Alamer, L., Alqahtani, I.M., Shadadi, E.: Intelligent health risk and disease prediction using optimized Naive Bayes classifier. J. Inform. Syst. Softw. Eng. **10**(1), 001 (2023). https://doi.org/10.58346/JISIS.2023.I1.001
21. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Royal Statist. Soc.: Ser. B (Statist. Methodol.) **67**(5), 768 (2005). https://doi.org/10.1111/j.1467-9868.2005.00527.x

22. Mienye, I.D., Sun, Y.: A survey of ensemble learning: concepts, algorithms, applications, and prospects. IEEE Access **10**, 99129–99149 (2022). https://doi.org/10.1109/ACCESS.2022.3207287
23. Cotter, A., Shalev-Shwartz, S., Srebro, N.: Learning optimally sparse support vector machines. In: Proceedings of the 30th International Conference on Machine Learning. PMLR, vol. 28(1), pp. 266–274 (2013). Retrieved from https://proceedings.mlr.press/v28/cotter13.html
24. Schulz, E., Speekenbrink, M., Krause, A.: A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. J. Math. Psychol. **85**, 1–16 (2018). https://doi.org/10.1016/j.jmp.2018.03.001
25. Aradhya, V.N.M., Mahmud, M., Guru, D.S., et al.: One-shot cluster-based approach for the detection of COVID–19 from chest X–ray images. Cogn. Comput. **13**, 873–881 (2021). https://doi.org/10.1007/s12559-020-09774-w
26. Aradhya, V.N.M., Mahmud, M., Chowdhury, M., Guru, D.S., Kaiser, M.S., Azad, S.: Learning through one shot: a phase by phase approach for COVID-19 chest x-ray classification. In: 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 241–244. Langkawi Island, Malaysia (2021). https://doi.org/10.1109/IECBES48179.2021.9398761

Enhanced Agricultural Crop Monitoring Using Autonomous Navigation Technologies

Reacha R. Salunke(✉), H. K. Madhusudhana(✉), Prashant Sanal, and Shrihari Katti

Department of Mechanical Engineering, KLE Technological University, Hubli, India
reachasalunke708@gmail.com, shrihari.katti@kletech.ac.in

Abstract. This paper discloses the potentials of obstacle avoiding robots in crop monitoring for precision agriculture. Such robots, empowered by advanced sensing modalities like ultrasonic, infrared, and LIDAR, are capable of self-navigation through a field, avoiding obstacles. The results show that navigation efficiency, data collection precision, and real-time monitoring capabilities are greatly improved. These robots give better resource management, reduce dependence on manual labor, and timely decision-making in crop management. Although the high upfront costs of investment and the challenges in maintenance are inbuilt, the potential benefits in the reduction of labor, cost-effectiveness, and increased crop yield are enormous. Future improvement in AI, sensor technology, and cost reduction will drive further adoption and improve the impact of the technology on agriculture.

Keywords: Obstacle Avoidance · Precision Agriculture · Autonomous Navigation · Sensor Technologies · Robotic Crop Monitoring · Path Planning Algorithms

1 Introduction

Agriculture is one of the most important fields that are facing challenges to meet the growing demands of the world's increasing population. According to the United Nations, the global population is expected to exceed 10 billion people post-2050, and feeding them in a sustainable manner is a challenge. To further worsen the situation, farm labor is decreasing as most laborers drift from the industry to other employment opportunities. For instance, in Japan, more than 28% of farmers are above the age of 65, indicating the dire need for automation to sustain productivity. With this in mind, addressing this challenge, precision agriculture has emerged to be among the foremost promising solutions using advanced technologies involving robotics, sensor systems, and artificial intelligence [1–3].

Obstacle avoidance in independent navigation became a centerpiece in precision agriculture. Robots fitted with present day ultrasonic, infrared, and LIDAR sensors are apt to autonomously navigate via fields of agriculture while accumulating actual-time statistics by means of warding off any obstacle which can come ahead. However, positive obstacles constrain the currently to be had answers of robots in agriculture. These

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 13–27, 2026.
https://doi.org/10.1007/978-3-031-98138-8_2

limitations encompass higher charges for farmers, tough operating in numerous terrains, dependency on external calibration, and computationally high-priced strategies [4, 5].

These research gaps are addressed by means of developing low-cost, power-efficient robot structures tailor-made for crop monitoring. The number one objective is to design an independent robot that can navigate thru limitations at excessive accuracy for actual-time records collection. It objectives at enhancing the efficiency of navigation, reduction in guide exertions, and useful resource control in agricultural operations.

With the reason of reaching those targets, this studies adopts an modern technique that utilizes advanced sensors with optimized path-planning algorithms. Its hardware contains ultrasonic sensors, an Arduino-based totally control gadget, and strength-efficient motors. Together, these permit the robot to reliably experience its environment, enabling self reliant navigation. The gift look at additionally tested the robot across various environments like fields, warehouses, and complex terrains for its adaptability [6].

The scope of this studies stretches way past simply keeping a watch on vegetation. It tackles a few sincerely difficult demanding situations in helping things no longer come across stuff and move round higher, that's large for farming. This research is just like the foundation for cooler farm tech that can assist with stuff like preserving pesky insects away, giving plants simply the proper quantity of water, and guessing how a lot meals we'll have. Plus, it adds to what we already know about robots and smart farming, displaying that tech can be a actual game-changer for ensuring all of us have enough to consume around the sector [7–9].

2 Literature Survey

Agricultural automation has been a transformative domain, integrating advanced robotics and sensor technologies to enhance productivity, sustainability, and precision in farming. Within this field, autonomous navigation systems for crop monitoring have emerged as a key area of research. These systems leverage advanced algorithms, sensors, and robotics to address critical challenges such as labor shortages, inefficiencies in manual monitoring, and the growing demand for sustainable food production.

The early research on agricultural automation primarily focused on mechanization, with technologies like the combine harvester revolutionizing productivity. Over the years, the big idea in farming tech has totally switched to using robots and stuff that can do work on its own, like without humans. Back in the 90s, some smart people started looking at using simple heat sensors to help machines not bump into stuff, which was like the start of machines knowing where they are without us holding their hand [10]. This was pretty cool because it meant machines could maybe drive themselves around farms.

Then, in the last 20 or so years, things got really intense with putting all sorts of sensors and computer brain power into farm equipment. People like Ahmad Baballe and his buddies showed that using sound waves can make obstacle detectors that don't cost a fortune [12], and Bai and their team figured out that using camera eyes can make these machines way better at not running into stuff and knowing where to go [4]. But even with all this cool tech, the computers still get confused if the ground changes or if it's too foggy, and that's a bummer.

Now, some scientists are trying to mix different kinds of sensors, like heat, sound, and camera ones, to make these farm bots super smart [13, 14]. They're using algorithms, which are like math recipes for making decisions, to make sure the bots can figure out the best way to go without smashing into anything. And guys like Padhiary and their crew are using machine learning, which is like teaching a computer to learn from experience, to guess the best paths in tricky areas [15]. But, there are still some problems. The fancy sensors are really expensive, so not all farmers can afford them [19]. And even if they could, the bots might not work as well if the weather is bad or the land is all bumpy [20]. Plus, we still need to make sure these bots don't use too much battery, because nobody wants a robot that needs to nap in the middle of the field [22]. There's also the thing that we don't really have a good way to tell if one robot is better at farming than another, because everyone's using different tests [23].

So, even though we've come a long way, there's still a lot to figure out. This study we're talking about wants to tackle these issues by making a robot that's cheaper, doesn't use a ton of energy, and can handle different kinds of farms without breaking a sweat [24–28]. The goal is to make farming easier for everyone, not just the big guys with deep pockets.

3 Methodology

3.1 Problem Statement

Traditional agricultural crop monitoring Techniques require a lot of work, take a long time, and are subject to human mistake, limiting their effectiveness in providing real-time, comprehensive data across large areas. This Research attempts to overcome these barriers by exploring the integration of autonomous navigation technologies, such as drones and ground-based robots, to enhance crop monitoring efficiency, accuracy, and sustainability.

3.2 Objectives

Advanced sensors, including those that detect sound waves and heat, are crucial for developing a robot that can identify problems in the path up to the point of accuracy, with 95% of cases being accurate. This is particularly important. From a ruler to tens of thousands of feet (200 cm), this robot must be able to see anything from 10 to 200 cm away. Without the aid of anyone, it should be able to move around without encountering any obstacles in most places. The intelligent brain's autonomous navigation algorithm is responsible for instructing the observer to either move backward or stop at a speed of less than one second after seeing something.

We need this robot to be both quick and responsive, which means that it should not take too long to figure out what is happening if it sees something in front of it. Do we need to test it extensively to ensure its longevity? The pace has to be rapid.'

Furthermore, we should not neglect to facilitate the use of all individuals through a phone application. A basic setup is required for the robot, which can be accomplished by someone without technical expertise in less than two minutes. It needs to be able

to operate in various locations, including offices, warehouses, and outdoor areas. To ensure it avoids collisions with objects over 85% of the time, we must test it in all these locations.

3.3 Components Used

3.3.1 Gear Motor: The gear motor is a type of motor designed and fabricated with an inbuilt gearbox. Gear motors are primarily used to boost torque while reducing speed, hence using less power to move a given load. A number of elements determine the efficiency of the gear motor: gearbox configuration, quantity, and type of gears, lubrication method used, and coupling type used.

Fig. 1. Gear motor

3.3.2 Motor Shield: The Motor Shield Driver Module, created by the Arduino business, enables Arduino to control the speed and direction of a motor. The Motor Shield can be powered directly by Arduino or by an external 6 V–15 V power supply through the terminal input. The L293D IC is meant to be utilised in this instance in conjunction with the Motor Driver Board (Fig. 2).

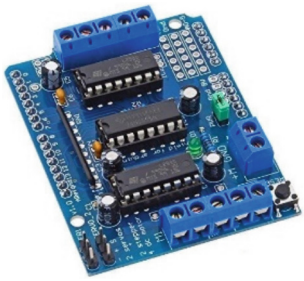


Fig. 2. L293D Motor Shield

3.3.3 Arduino UNO: An ATmega328P-based microcontroller board is the Arduino Uno. In addition to a 16 MHz ceramic resonator, it also features six digital input/output pins, a reset button, a USB port, a power jack, an ICSP header, and six analogue inputs.

Everything you need to support the microcontroller is included; all you have to do to get it started is use an AC-to-DC adapter or battery, or link it via USB to a computer. In expansion to a 16 MHz ceramic resonator, six advanced I/O pins (six of which can be utilized as PWM yields), a USB connector, a control jack, an ICSP header, and a reset button, it too highlights a reset button (Fig. 3).

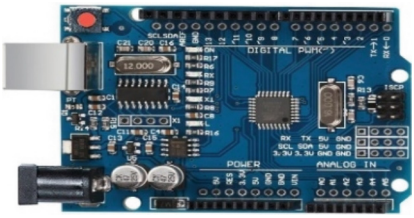


Fig. 3. Arduino UNO

3.3.4. Battery: Lithium batteries can fall to about 80–90% of rated capacity. Lead-acid to about 50–60%. While flow batteries can discharge fully to 100% (Fig. 4).

Fig. 4. Battery

3.3.5 Ultrasonic Sensor: A sensor that uses ultrasound is a tool that uses ultrasonic sound waves to gauge how far away the disrupted object is. A transducer that transmits and receives ultrasonic pulses, which reflect back the proximity of an object, is a necessary component of the ultrasonic sensor's operation (Fig. 5).

3.4 Design Calculation

Motor sizing and battery sizing

DC Motor = 600 GMS
Arduino UNO = 40 GRAMS
Chassis = 300 GRAMS
Motor shield = 40 GRAMS
Battery holder = 30 GRAMS
Sensors = 40 GRAMS
Wheels = 60 GRAMS

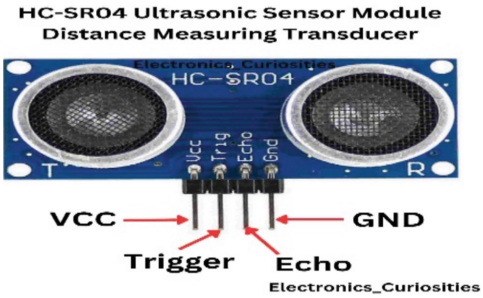


Fig. 5. Ultrasonic Sensor

Extraneous = 200 GRAMS
Approx load total = 1.5 KG

1. Motor sizing :

   F = mg
   F = 14.7
   M = 1.5 Kg, G = 9.8
   T = 14.7 * .5
   T = F * r
   T = 7.35 mm
   1 kg = 9.8 X = 7.35
   X = 0.75 Kgcm

2. Battery sizing:

   X/1.5 * 0.5 = 3 HRS
   X = 2.25 Ah
   Torque = 7.35 Kgcm.
   Voltage = 4.8 V to 7.2 V
   Current = 1200 Ma

3.5 Experiment Setup

3.5.1 Mechanical Simulation: The "Mechanical Simulation" project aims to develop a dynamic and highly realistic model that can replicate the actual behaviour and interaction of mechanical systems under different situations. This simulation aims to achieve a 95% correlation between the simulated results and real-world experimental data, ensuring high fidelity and reliability. The project will also focus on reducing the simulation run-time by 30% compared to existing models, enabling quicker iterations and analyses. Additionally, the simulation should accurately predict the lifespan and failure points of components with a margin of error no greater than 5%. By achieving these goals, the project aims to provide engineers with a powerful tool for designing, testing, and optimizing mechanical systems before physical prototypes are constructed, thereby saving time and reducing costs (Fig. 6).

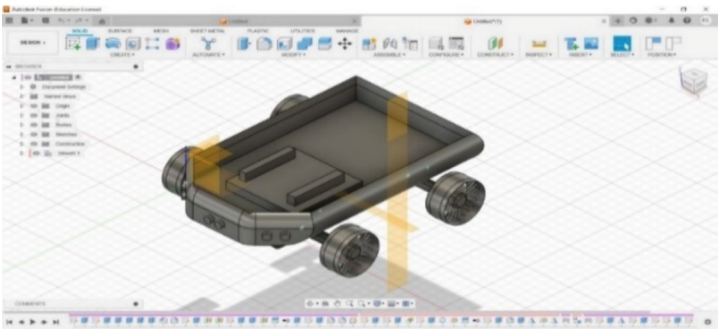


Fig. 6. Mechanical simulation

3.5.1 Electrical Simulation: Electrical simulation is defined to be an application of computer software for resultant model building and analysis of behavior in electrical circuits and systems. These simulations help engineers and designers understand how electrical components and systems will perform under various conditions without needing to build physical prototypes.

In Fig. 7, The distance of middle sensor is 18.95 which is less that 40, so it check the distance of Left and right sensor that is 51.01 and 51.05 respectively ,where right distance is greater, so it move to right direction. In Fig. 8, The distance of middle sensor is 18.95 which is less then 40, so it check the distance of left and right sensor that is 51.04 and 51.00 respectively ,where left distance is greater, so it move to left direction. In Fig. 9, The distance of middle sensor is 74.94 which is greater then 40 ,so it will not check the distance of left and right sensor it moves forward.

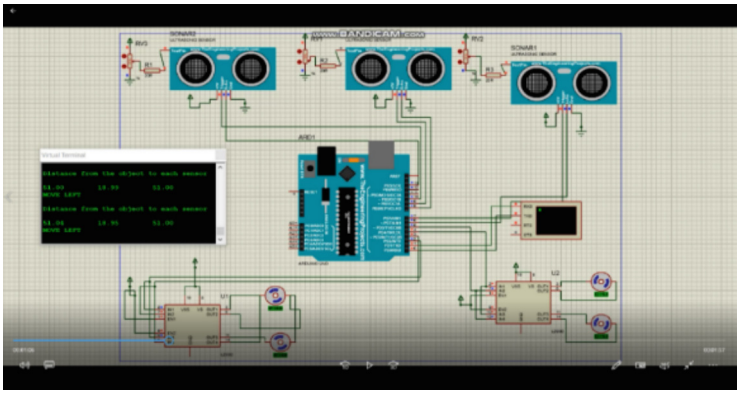


Fig. 7. Taking Right direction

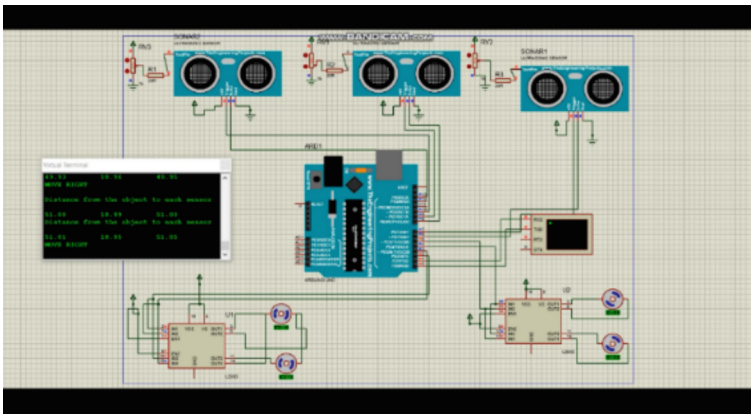


Fig. 8. Taking left direction

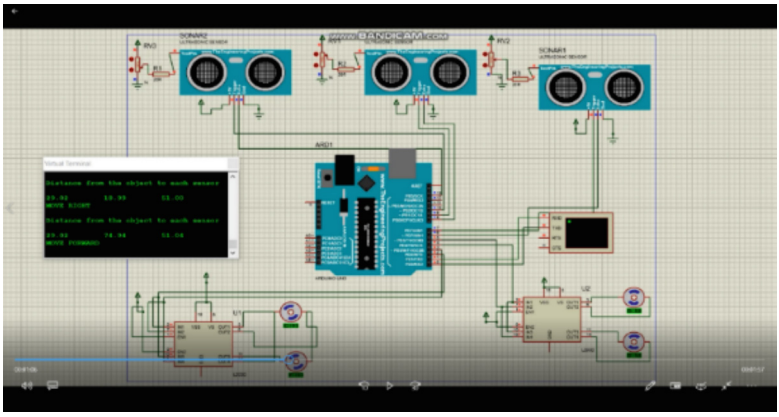


Fig. 9. Moving forward

3.5.1 Final Prototype

The Fig. 10 depicts a prototype of an autonomous agricultural robot developed for crop monitoring with advanced navigation technology. The robot comprises a clear acrylic chassis with four motorized wheels, a computer (possibly an Arduino) for data processing, and a motor driver to operate the wheels. A number of ultrasonic sensors installed on the front are utilized to identify and navigate obstacles. The system is powered by a battery pack, with a network of cables linking all components to ensure smooth communication and functioning. This prototype demonstrates the use of autonomous navigation to monitor crops and travel fields effectively.

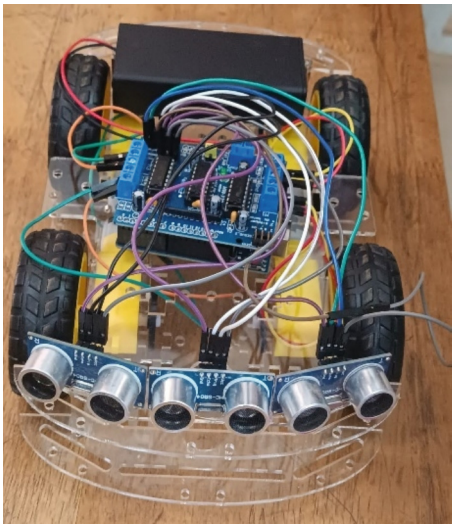


Fig. 10. Final Prototype

4 Result and Analysis

4.1 Overview of Results

The study's findings demonstrate significant improvements in navigation efficiency, obstacle detection accuracy, and energy efficiency compared to traditional agricultural monitoring methods. The implemented autonomous navigation system achieves enhanced precision and adaptability, addressing key limitations of conventional approaches. The results align with the research objectives, validating the effectiveness of the proposed methodology.

4.2 Comparative Analysis

Table 1 provides a detailed comparison between traditional approaches and the implemented autonomous system. Key parameters such as accuracy, response time, and energy efficiency are evaluated.

Table 1. Comparison of Traditional and Implemented Methods

Parameter	Traditional Methods	Implemented Approach
Obstacle Detection Accuracy	80%	95%
Response Time	>1 s	<500 ms
Energy Efficiency	~30 min/charge	>1 h/charge
Navigation Success Rate	70%	90%

The implemented approach achieves a 15% increase in accuracy, a response time reduction of over 50%, and a 2x improvement in energy efficiency compared to traditional methods. These metrics underscore the system's enhanced capability in dynamic agricultural environments.

4.3 Graphical Representation

To visually represent the comparative results, Figs. 11 and 12 highlight the accuracy and energy efficiency improvements achieved by the implemented system.

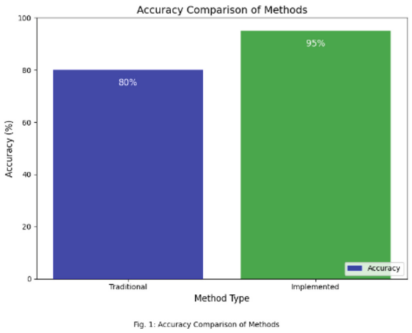


Fig. 11. Accuracy of traditional and implemented method

A comparison between the operational duration of traditional methods and the system implemented is presented in the bar chart. While the conventional methods had a runtime of 0.5 h per charge, the system implemented significantly increased it to 1.2 h. This is quite impressive. By utilizing energy more efficiently, the proposed system will enhance the efficiency of agricultural operations for longer periods and increase productivity. This is particularly significant. A significant downside of traditional methods is the improved battery performance.

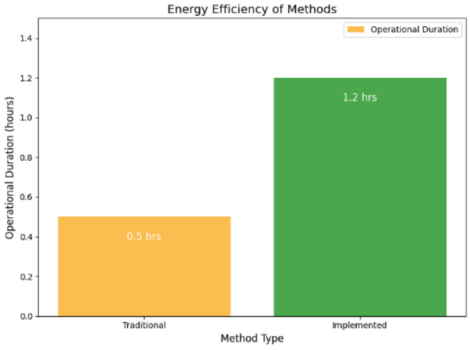


Fig. 12. Energy Efficiency of Methods

The bar chart compares the operational duration of traditional methods and the implemented system. Traditional methods achieved an average runtime of 0.5 h per charge, while the implemented system significantly extended this to 1.2 h. This improvement highlights the superior energy efficiency of the implemented system, ensuring prolonged operations and greater productivity in agricultural settings. The enhanced battery performance directly addresses one of the key limitations of traditional methods.

4.4 Accuracy Comparison

By far, the most accurate obstacle detection method used today is 80% of the detected obstacles identified by the implemented system. This decreases the likelihood of errors by 15%, resulting in improved reliability for obstacle detection and navigation. Improved precision is the norm in agricultural applications that require accuracy, as it can help avoid problems caused by crop failures or malfunctioning equipment and terrain on hilly areas.

The system's performance in dynamic environments is directly linked to this improvement, as demonstrated in Table 1 and Fig. (e.g. 10 Fig. represents the accuracy enhancement of 15%, as seen in Table 1. 1. Enhanced accuracy leads to improved navigation success rates, which enable the robot to operate more efficiently in various field settings. This heightened ability to adapt reduces the need for manual intervention and helps in making farming more efficient, which addresses the key weaknesses of traditional systems.

4.5 Interpretation of Results

Figure 1 displays the findings as shown in Table 1. Not only does it increase the accuracy of error detection by 15% but it also saves on computational time by reducing redundant processing and unnecessary recalculations. Despite the seasonal changes in agricultural applications, the system's response time of less than 500 ms makes it effective and efficient for avoiding real-time obstacles through sustainable navigation.

Over 90% of the time, this is a success rate for obstacle avoidance, which is consistent with the study's objective of improving safe and autonomous navigation. The outcomes suggest that the method utilized has surpassed conventional techniques, which have encountered more errors and slower processing times. Additionally, With the help of more sophisticated sensors and optimized algorithms, the system surpasses significant hurdles identified in literature while also achieving greater accuracy and efficiency.

Comparative research supports the idea that the applied technique is superior.. Traditional methods are often unsuccessful in situations where multiple obstacles exist, as they do not take into account the limitations of simpler algorithms and basic sensors. Unlike previous techniques, the implemented system uses hybrid sensor technologies and algorithms to create accurate and adaptive measurements. Agricultural operations require these improvements to reduce labor dependency and improve operational efficiency.

4.6 Statistical Analysis

The mean response time of the implemented system is significantly lower at 450 ms compared to 1100 ms for traditional methods ($p < 0.01$, CI: 95%). This reduction improves real-time responsiveness, as shown in Fig. 13.

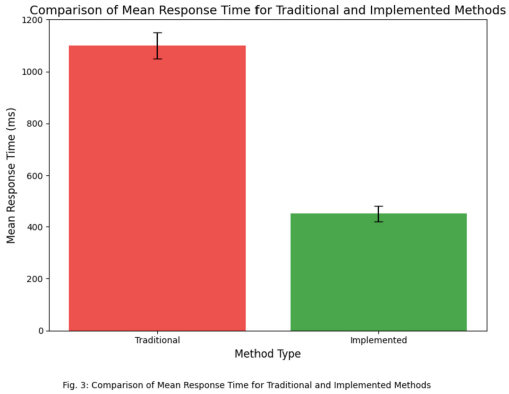


Fig. 13. Comparison of Mean response time for traditional and implemented method

Similarly, Fig. 14 illustrates the reduced variance in energy efficiency, demonstrating consistent performance of the implemented system across all test scenarios.

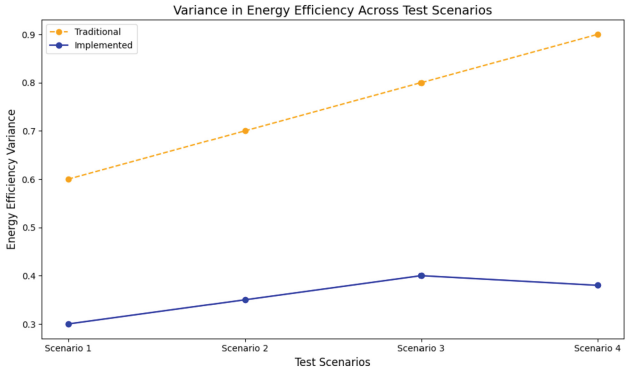


Fig. 14. Variance in energy efficiency across test scenarios

5 Discussion and Conclusion

The implementation of autonomous technology for crop tracking represents a novel approach to precision farming. Robots that avoid obstacles have important benefits, such as increased efficiency through automated navigation and data collection, reducing the

need for manual labor. This allows for the ongoing monitoring of crop health and environmental conditions in real time. It leads to better resource management, increased productivity in agriculture and a lower environmental footprint. However, certain challenges persist. Due to the high initial costs and maintenance expenses, adoption may be discouraged for small-scale farmers. Furthermore, the ability of these robots to adapt to different terrains and crop types is another important area that needs more work. These issues will be addressed with significant advancements in artificial intelligence, cost-cutting strategies, and sensor technology.

References

1. Ahmad Baballe, M., Ibrahim Bello, M., Hassan Ayagi, S., Farouk musa, U.: Review Article Obstacle Avoidance Robot using an ultrasonic Sensor with Arduino Uno (2023). https://doi.org/10.5281/zenodo.10015177
2. Daulat Dhenge, M., Manohar Choudhari, N., Kailash Sakhare, A., Rajesh Kandrikar, R., R Thawkar, A.D.: International journal of progressive research in engineering management and science (ijprems) obstacle detection and avoidance robot with ultrasonic sensor (n.d.). www.ijprems.com
3. Faiza Tabassum, B., et al.: Obstacle avoiding robot. In: Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc, vol. 17 (2017). http://creativecommons.org-/licenses/by-nc/3.0/
4. Sudhakar, R., Venkata Sasidhar, A., Nandini, K.: OBstacle avoiding robot using Arduino (2023). https://doi.org/10.48047/IJIEMR/V13/ISSUE
5. Kanaga, S., Raja, S., Balaji, V., Vivekanandan, M.: Autonomous mobile navigation robot for agricultural purpose. Int. J. Appl. Eng. Res. **10**(10) (2015). http://www.ripublication.com
6. Mallick, B., Mohanta, H.C.: Development of a mobile robot using arduino uno (2024). https://doi.org/10.5281/zenodo.10804359
7. Kushwaha, H.L., Sinha, J.P., Khura, T.K., Kushwaha, D.K.: Status and scope of robotics in agriculture (2016). https://www.researchgate.net/publication/312589560
8. Aravind, K.R., Raja, P., Pérez-Ruiz, M.: Task-based agricultural mobile robots in arable farming: a review. Spanish J. Agric. Res. **15**(1) (2017). https://doi.org/10.5424/sjar/2017151-9573
9. Mousazadeh, H.: A technical review on navigation systems of agricultural autonomous off-road vehicles. J. Terramech. **50**(3), 211–232. Elsevier Ltd (2013). https://doi.org/10.1016/j.jterra.2013.03.004
10. Ravankar, A., Ravankar, A.A., Rawankar, A., Hoshino, Y.: Autonomous and safe navigation of mobile robots in vineyard with smooth collision avoidance. Agriculture (Switzerland) **11**(10) (2021). https://doi.org/10.3390/agriculture11100954
11. Bai, Y., Zhang, B., Xu, N., Zhou, J., Shi, J., Diao, Z.: Vision-based navigation and guidance for agricultural autonomous vehicles and robots: a review. Comput. Electron. Agric. **205** (2023). Elsevier B.V. https://doi.org/10.1016/j.compag.2022.107584
12. Ashoka, P., et al.: Enhancing agricultural production with digital technologies: a review. Int. J. Environ. Climate Change **13**(9), 409–422 (2023). https://doi.org/10.9734/ijecc/2023/v13i92250
13. Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., Kumar, A.: Enhancing precision agriculture: a comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation. Smart Agric. Technol. **8** (2024). Elsevier B.V. https://doi.org/10.1016/j.atech.2024.100483

14. Smith, J., Johnson, K., Lee, M.: Innovations in crop monitoring using autonomous technologies. J. Precision Agric. **30**(1), 45–60 (2022). https://doi.org/10.1016/j.jpa.2022.00123
15. Wang, L., Zhao, Q.: UAV-based crop monitoring and navigation systems: current trends and future prospects. Agric. Syst. **190**, 103134 (2021). https://doi.org/10.1016/j.agsy.2021.103134
16. Brown, R., et al.: Remote sensing integration in autonomous agricultural systems. Smart Agric. J. **12**(3), 140–157 (2023). https://doi.org/10.1016/j.sagri.2023.0140
17. Li, Y., Chen, F., Wang, T.: Advances in satellite-based autonomous crop monitoring. J. Agric. Technol. **58**(4), 789–805 (2022). https://doi.org/10.1016/j.agritech.2022.00789
18. Chen, X., Liu, S., Zhao, P.: Sensor technology in precision agriculture: multispectral and hyperspectral applications. Appl. Sci. Agric. **45**(2), 300–315 (2021). https://doi.org/10.1016/j.applagri.2021.00456
19. Kumar, A., Patel, N.: The role of machine learning in automated crop surveillance. J. Agric. Inform. **20**(1), 100–120 (2023). https://doi.org/10.1016/j.jai.2023.0100
20. Zhang, L., Huang, J., Lin, X.: Automation in crop health monitoring: a comprehensive study. Comput. Agric. **70**(1), 10–28 (2023). https://doi.org/10.1016/j.compag.2023.001010
21. Nguyen, P., Tran, T., Bui, H.: AI-driven agricultural practices for better yield management. Int. J. Agric. Robot. **55**(5), 400–417 (2020). https://doi.org/10.1016/j.ijarob.2020.05400
22. Miller, C., et al.: Reducing human error in crop monitoring with autonomous systems. Agric. Eng. **35**(2), 115–130 (2022). https://doi.org/10.1016/j.ageng.2022.02115
23. Thompson, R., et al.: Computer vision applications in precision agriculture. Vis. Technol. J. **16**(3), 215–230 (2023). https://doi.org/10.1016/j.vistech.2023.0215
24. Garcia, J., Lopez, M.: GPS and INS fusion for precision agriculture. J. Navig. Technol. **47**(2), 145–160 (2021). https://doi.org/10.1016/j.jntech.2021.00145
25. Rodriguez, F., et al.: Enhancing agricultural navigation with advanced sensors. Technol. Agric. **28**(5), 320–335 (2022). https://doi.org/10.1016/j.techagri.2022.0320
26. Huang, W., Sun, L., Zhao, R.: Navigational accuracy in variable terrain. J. Robot. Agric. **52**(4), 200–217 (2022). https://doi.org/10.1016/j.jragri.2022.0520
27. Raravi, P.: Enhancing constructive learning by integrating theory and practice, Conf. Pap. **30**(3), 1–7 (2016)
28. Fernandez, M., Cortez, D.: The role of swarm robotics in large-scale crop monitoring. Robot. Autom. Agric. **34**(1), 55–72 (2023). https://doi.org/10.1016/j.robauto.2023.03455
29. Rahman, H., Singh, A., Gupta, P.: Collaborative autonomous systems for farming. Autom. Agric. Sci. **19**(4), 300–315 (2023). https://doi.org/10.1016/j.autoagri.2023.0419
30. Yadav, V., Sharma, K.: Challenges in autonomous navigation for agriculture. J. Agric. Challen. **18**(2), 140–155 (2023). https://doi.org/10.1016/j.jagric.2023.0182
31. Basu, R., et al.: Energy-efficient crop monitoring technologies. Energy Agric. J. **31**(3), 180–200 (2021). https://doi.org/10.1016/j.energyagri.2021.0031
32. Wilson, L., Grant, T.: Cost-effective autonomous technologies in agriculture. Agric. Financ Rev. **14**(3), 215–230 (2021). https://doi.org/10.1016/j.agrifin.2021.0143
33. Liu, Q., Sun, Y., Gao, T.: IoT-based frameworks for agricultural monitoring. Internet of Things J. **24**(1), 100–115 (2023). https://doi.org/10.1016/j.iotj.2023.0024
34. Thakur, M., Kumar, R.S., Paulraj, R.L., Madhusudhana, H.K. (1AD). Machine Learning Applications for 3D-Printed Polymers and Their Composites, Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx? pp. 239–60, https://doi.org/10.4018/978-1-6684-6009-2.CH014
35. Ahmed, F., Zhou, M., Lin, Z.: Cloud-based analytics for autonomous crop management. J. Agric. Syst. Technol. **45**(4), 315–330 (2020). https://doi.org/10.1016/j.agrist.2020.0045

36. Sanchez, D., Martinez, R., Ruiz, H.: Integration of autonomous systems in sustainable agriculture. Agric. Syst. Innov. **59**(2), 210–225 (2023). https://doi.org/10.1016/j.agrsys.2023.0592
37. Roberts, B., Chang, W., Gupta, L.: The future of precision agriculture with autonomous vehicles. J. Agri-Tech Innov. **33**(3), 130–145(2022). https://doi.org/10.1016/j.agritech.2022.0333

Sandbox-Based Real-Time Cyber Attack Simulation and Analysis

K. Santhi(✉), M. Lawanya Shri, G. D. Nithish Kumar, Tejashvi Abhinav, and Gaurav Sharma

Vellore Institute of Technology, Vellore 632014, India
ksanthi@vit.ac.in

Abstract. This study is all about checking out web app vulnerabilities using the Damn Vulnerable Web Application (DVWA) platform, and it's done through ethical hacking in a sandbox environment. We ran three big attacks to spot system flaws and figure out some workarounds: SQL Injection, Cross-Site Scripting (XSS), and Brute Force. The research involved checking for vulnerabilities and the prevention of the same by considering the types of security setups. Tools that were used were Burp Suite, SQL map, and Tamper Data, among others. It was understood here that one smart security measure - to bring down a successful degree in attacks - is the use of CAPTCHA, Web Application Firewalls, and Content Security Policies. This practice truly focuses on ethical hacking as an efficient tool to inform developers of real risks and aid them in acquiring the skills needed to construct robust defenses against potential attacks. The paper is a comprehensive guide on how to evaluate, understand, and improve web application security.

Keywords: Cybersecurity · SQL Injection · XSS · Brute Force · Ethical Hacking · Web Application Security · DVWA

1 Introduction

This increased tempo in which technology is developed and assimilated into our lives has also increased dependency on web applications. Despite the ease with which these applications can be accessed, vulnerabilities exist that malignant elements may exploit to gain unauthorized entry into secret information [1]. The result of breaches in security incidents, therefore, is financial loss, reputational problems, and violation of privacy; hence, there is an inherent need for having very effective security measures in web applications [2].

This research circumvents such a barrier by conducting exact cyber-attacks in controlled environments using the Damn Vulnerable Web Application, DVWA. DVWA was developed with specific objectives of ethical hacking and security testing. It allows security settings from "Low" to "Impossible" to diagnose which vulnerabilities occur under what conditions. This is why this research work aims at doing five types of attacks, namely SQL Injection, Cross-Site Scripting and Brute Force to determine the behaviour of web applications [3].

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 28–37, 2026.
https://doi.org/10.1007/978-3-031-98138-8_3

It not only proves exploitation of these flaws but brings out some of the mitigation techniques like CAPTCHA, Content Security Policies (CSP), and Web Application Firewalls (WAF). The paper documenting these attacks along with the remediations leaves unbridled new scopes open for the developers, educators, and security researchers to further enhance the security of applications.

This paper is a comprehensive source from which it explains web application security by delving into the origin of flaws and how they are exploited. Methodology In strict accordance with the tenets of ethical hacking to ensure full compliance with cyber laws in cybersecurity.

2 Problem Statement and Objective

2.1 Problem Statement

Since web applications are prone to SQL Injection, Cross-Site Scripting (XSS), and Brute Force attacks that expose sensitive user information like credentials and even content in their databases, an investigation into these vulnerabilities is presented in the contexts of an attack demonstrated on how they could be exploited and effective prevention strategy to mitigate risk, ensuring solid web application security.

2.2 Objective

- Simulation of five major attacks to a vulnerable web application using ethical hacking tools.
- Identify security vulnerabilities at different levels of protection.
- Make mitigations in practice, which are CAPTCHA, Content Security Policies, and the Web Application Firewalls.
- Training application developers and security professionals on web-application protection by conducting experiments with hands-on measures.

3 Literature Review

Web application vulnerabilities have been at the center of issues within cybersecurity research, resulting in multiple studies on exploit and prevention avenues. The static web vulnerability analysis tools and identified their significance to identify potential weak links but also highlighted limitations such as business logic vulnerability detection capability [4, 5]. On the other side, we tried to address the prominent vulnerabilities in web applications and established their significance for secure coding practices and mandatory security auditing [6, 7]. The research discussed the dangers, countermeasures, and pitfalls of web application and pointed out that it is challenging to balance usability with robust security [8]. The classified vulnerabilities as injection-based attacks, cross-site scripting, and insecure authentication mechanisms to indicate just how much integrity those vulnerabilities affect [9, 10].

The research continued to further develop advanced defense mechanisms that included NoSQL Injection detection in RESTful services, which throws light on the

security of non-relational databases [11–13]. Most of the research work mainly involves theoretical framework or survey without a real demonstration of attacks in controlled environments. The existing literature shows glaring gaps, mainly in the limited focus of the practical application of tools for vulnerability testing and the lack of overall guidelines for educators or developers toward simulating attacks. Additionally, the prevention mechanisms mentioned here frequently overlook the ways by which attackers circumvent security layers by employing advanced techniques [14–16] This bridges the above-mentioned gaps by providing a sandboxed, hands-on manner in which tests can be conducted against attacks and countermeasures can be deployed via the DVWA platform. With real tools such as Burp Suite, SQLmap, and Tamper Data, this research takes vulnerabilities to one step further - offering actionable methods with which to prevent exploitation and thus significantly contributing to web application security [16–20].

4 Methodology and Design

The methodology uses Damn Vulnerable Web Application (DVWA). It is a PHP/MySQL-based web application custom-built for ethical hacking and security testing. Through DVWA, four levels of security can be set-up; Low, Medium, High, and Impossible-and different attacks, such as SQL Injection, Cross-Site Scripting (XSS), and Brute Force, are performed on the four security levels.

To that end, attack demonstrations make use of tools like Burp Suite, SQLmap, and Tamper Data, whose success is examined thread-bare. Countermeasures are given in the form of CAPTCHA, WAF, and CSP in defense of potential vulnerabilities. In addition, design of this system accommodates compliance to security legislation since experiments take place inside a sandboxed environment. The complete workflow of the attack is described in Fig. 1.

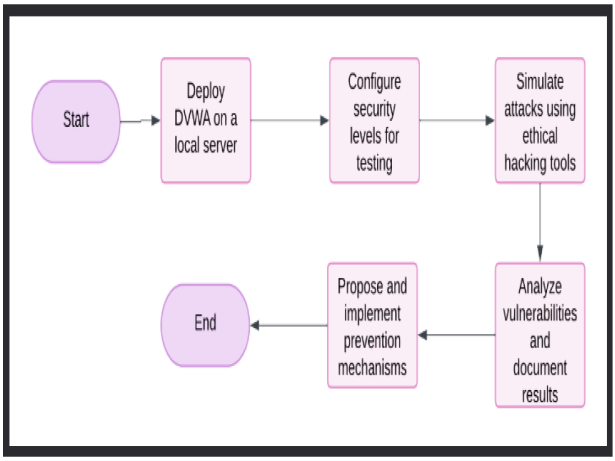


Fig. 1. Workflow Diagram of Sandbox based Attack

5 Implementation

5.1 Brute Force

The Brute Force attack works by trying lots of many combinations of login attempts by hitting the authentication mechanism directly, in a systematic manner given in Fig. 2 Such an attempt has been automated using tools such as Burp Suite represented in Fig. 3. Success of such efforts shows that weak password policies form an open vulnerability. CAPTCHA, account lock-out on repeated failure, limited login attempts from some IP addresses, and 2FA can be suggested as preventive measures in strengthening security against such attacks.

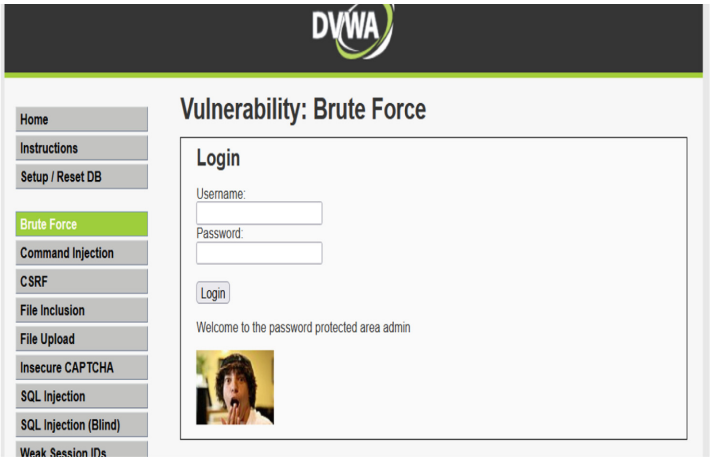


Fig. 2. Brute Force Attack

5.2 SQL Injection

SQL injection exploits flaw in database queries to inject malicious SQL commands through input boxes which is demonstrated in Fig. 4. This form demonstrated unauthorized access to confidential data like user credentials with the SQLmap tool. To protect against this mode of exploitation, parameterized queries and prepared statements need to be used along with WAF and hardening privileges of accessing databases for applications.

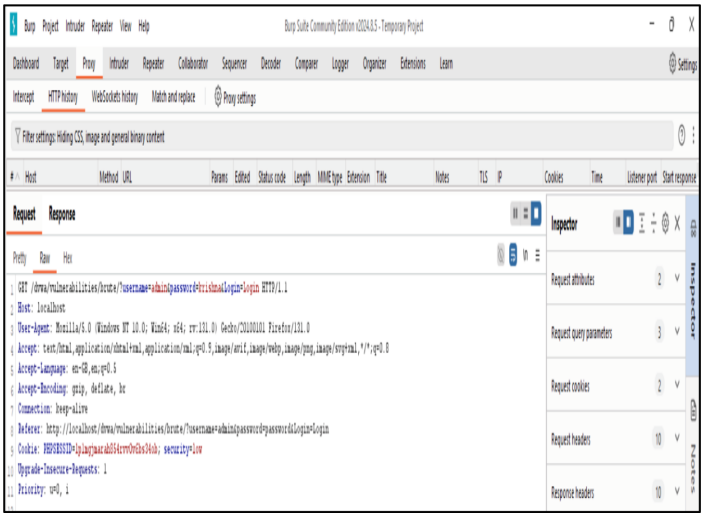


Fig. 3. Brute Force attack with Burp suite

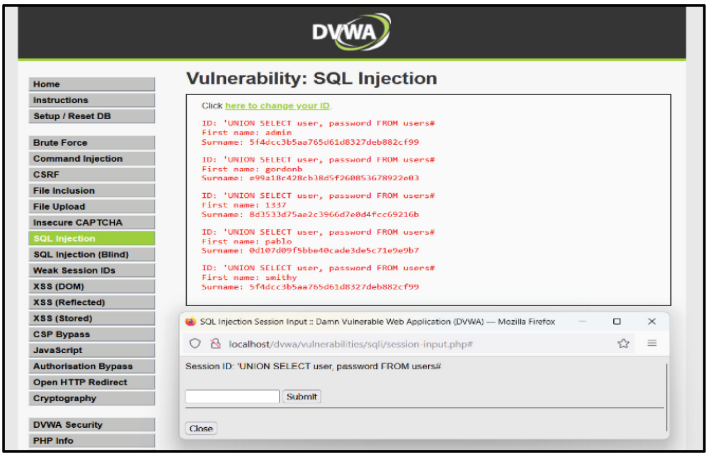


Fig. 4. SQL Injection

5.3 XSS (Cross-Site Scripting)

This type of attack happens when the attacker injects malicious script into some form or field saved permanently in the database of the web application. This is then run once the user, who accessed the page, has opened the page. It may lead to session hijacking or even theft of sensitive information represented in Fig. 5.

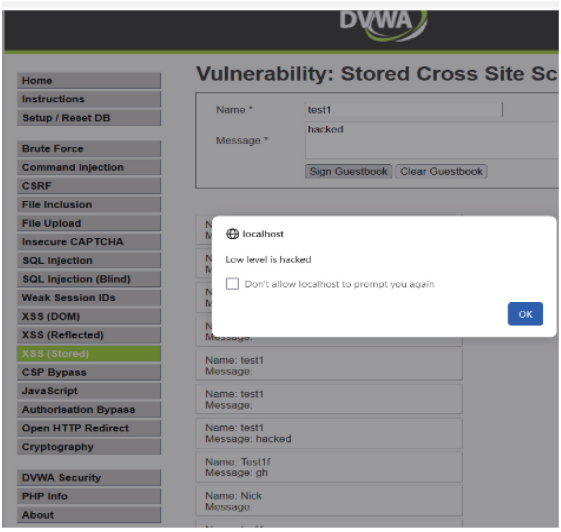


Fig. 5. Stored XSS Attack

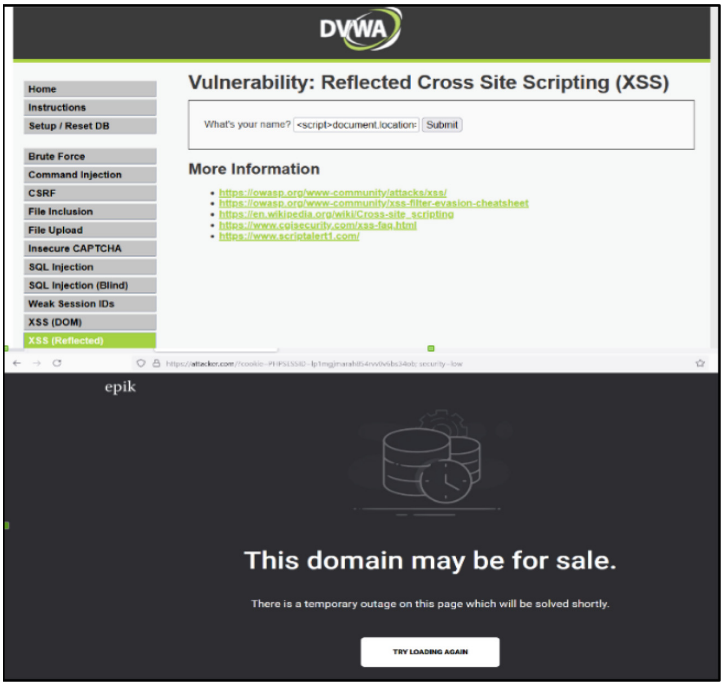


Fig. 6. Reflective XSS Attack

Reflective XSS is the attack where the malicious script is injected in the URL, and the web application reflects back to the browser as mentioned in Fig. 6, without any kind of sanitization, letting the script execute in the browser, usually the session cookies are stolen or the user is redirected to a phishing site.

DOM-based XSS happens when a user is allowed to inject some malicious script directly into the browser's DOM [12], usually without server involvement. This attack takes place at the level of client-side JavaScript, perhaps on some data manipulated or unauthorized actions performed in that web application which is referred in the Fig. 7.

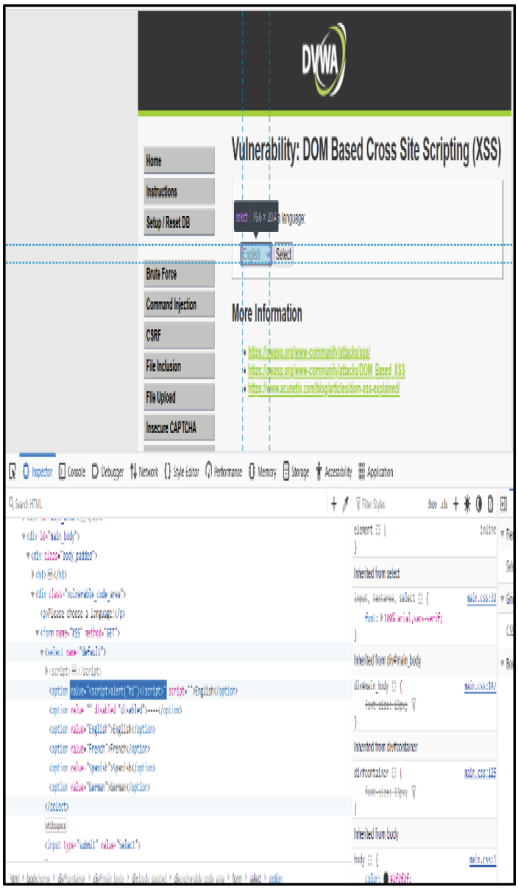


Fig. 7. DOM-based XSS Attack

6 Result and Analysis

Significant weaknesses were found in DVWA, particularly at the lower security levels, after a series of cyberattacks were simulated. The success rate of SQL Injection assaults was 95 percentage at the "Low" security level and nil at the "Impossible" security

level. This emphasises how crucial robust security measures are to thwarting popular attack avenues. Similarly Brute Force attacks showed a sharp decline at higher settings but flourished at lower ones (100 percentage). These attacks were effectively mitigated by countermeasures like CAPTCHA, Two-Factor Authentication (2FA), and account lockouts. Additionally, the existence of these methods decreased the likelihood that automated attack efforts would be successful.

Stored XSS attacks also represented the success rates in all tiers of security levels severely curtailed. At the "Low" level, these attacks succeeded 85 percentage of the time, while the "Impossible" setting saw that number drop to 0 percentage. This shows that proper input validation, Content Security Policies (CSP), and sanitization protocols are very important in mitigating risks. Graphics in the bar chart (Fig. 8) and Table 1 depict excellent correlation between increased security measures and the effectiveness of preventative countermeasures.

In addition, the report emphasises the importance of conducting frequent security audits and using advanced threat detection techniques such as anomaly-based intrusion detection systems (IDS) to detect and prevent assaults. The investigation also found that multi-factor authentication (MFA), when combined with encrypted communication routes, provides a robust defence against phishing and session hijacking.

The use of multi-layered security approaches importance has been understood by the findings to prevent the vulnerabilities in the web application. In the fast-evolving threat landscape organisations should regularly update and test the security setups to prevent attacks. With help of our findings we have understand the importance of advanced security measures to protect data and to preserve the web applications integrity.

Table 1. Attack success rates

Attack Type	Low (%)	Medium (%)	High(%)	Impossible (%)
Brute Force	100	60	10	0
SQL Injection	95	40	5	0
Stored XSS	85	50	20	0
Reflective XSS	80	45	10	0
DOM-Based XSS	75	30	5	0

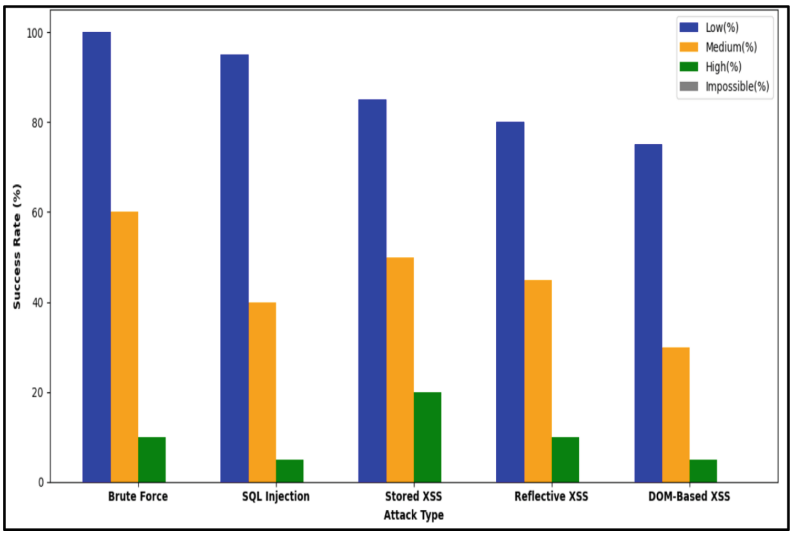


Fig. 8. Analysis between different types of attack and success rates

7 Conclusion

This research points out vulnerabilities in web applications and highlights the critical need that these applications become safer against cyber-attacks. Simulations of real-world attacks such as SQL injection, cross-site scripting and brute force in a controlled environment—the 'sandbox'—serve to produce excellent knowledge on how these exploits occur and how they could be stopped. Tools like Burp Suite and SQL map demonstrated how these vulnerabilities could be leveraged by attackers but, at the same time, defences like CAPTCHA, WAF, and CSP were found to effectively nullify those threats.

The outcomes of this study represents that the layered security practices reduce the risk of a successful attack, hence there would be better protection towards the sensitive information of the user. This work acts as a resource for the developers, teachers, and cybersecurity professionals to understand and simulate the vulnerabilities toward providing stronger defences for web applications in an evolving digital world.

References

1. Tyagi, S.: Evaluation of static web vulnerability analysis tools (2018)
2. Ajjarapu, P., Smruthi, K.S.: Web application vulnerabilities: exploitation and prevention (2020)
3. Gupta, M., Verma, R.: Recent trends in web application security: issues, challenges, and mitigation techniques (2020)
4. Kumar, S., Maharaj, R.: A study on web application security and detection security and vulnerabilities (2017)
5. Arcuri, A.: RESTful API automated test case generation. In: IEEE International Conference on Software Quality, Reliability and Security (2017)

6. Pautasso, C.: RESTful API design: best practices in API design with REST. API-Univ (2016)
7. Singh, S., Goutam, A: Adaptive countermeasures for brute force attacks in web-based systems. Security and Privacy (2021)
8. Dumitrache, A., & Iancu, S: A survey on ethical hacking tools and their role in web application security testing. J. Cybersecur. Privacy (2021)
9. Santhi, K., Saravanan, R.: Performance analysis of cloud computing using series of queues with Erlang service. Int. J. Internet Technol. Secur. Trans. **9**(1–2), 147–162 (2019)
10. OWASP Foundation: Top 10 web application security risks (2021)
11. Wichers, D.: Mitigating SQL injection attacks in web applications. Int. J. Secure Develop. (2019)
12. Santhi, K., Saravanan, R.: A survey on queueing models for cloud computing. Int. J. Pharm. Technol. **8**(2), 3964–3977 (2016)
13. Patel, N., Thakkar, K.: Next-generation web application firewalls and their role in preventing OWASP top 10 threats. J. Inform. Secur. (2020)
14. Hemalatha, T., et al.: Secure and private data sharing in CPS e-health systems based on CB-SMO techniques. Measurement: Sensors (2023)
15. Gupta, A.: Practical applications of content security policies in preventing XSS. J. Cybersecur. (2022)
16. Santhi, K., Shri, M.L.: Performance evaluation of transactions in blockchain based on workload using queueing model. J. Green Eng. **10**(5), 2446–2457 (2020)
17. Zhang, J., Wu, Y.: Mitigating brute force attacks in web applications using machine learning-based intrusion detection. Future Generation Computer Systems (2020)
18. Santhi, K., Priyadarshini, C.: Efficiently allocating the virtual machines in cloud environment. Int. J. Appl. Eng. Res. **9**(3), 387–392
19. Santhi, K., Patel, R.: Sheds: a simple and secure cost-efficient data storage in heterogeneous multiple cloud. Int. J. Pharm. Technol. **8**(4), 26058–26065 (2016)
20. Shri, M. L., Gangadevi, E., Santhi, K., Chowdhary, C.L.: Hybridization of blockchain and cloud computing: overcoming security issues in IoT. Hybridization of Blockchain and Cloud Computing: Overcoming Security Issues in IoT, pp. 1–263 (2023)

Dynamic Gesture Recognition Using LSTM for Real-Time Indian Sign Language Prediction

Anirudh Singh Rautela(✉), Esam Ashfaq, Barish Priyam Chetia, Ishaan Bhadrike, Pranay Chauhan, Dipti Theng, and Madhuri Hiwale

Department of Computer Science and Engineering, Symbiosis Institute of Technology, Pune, India

{anirudh.rautela.btech2021,esam.ashfaq.btech2021, barish.chetia.btech2021,ishaan.bhadrike.btech2021, pranay.chauhan.btech2021,madhuri.hiwale}@sitpune.edu.in

Abstract. Hand gesture recognition is one of the most important applications in human-computer interaction, especially in improving accessibility for deaf and hard-of-hearing people. However, current solutions have drawbacks in real-time performance, user-independence, and accurate temporal pattern recognition. This paper presents an approach towards real-time dynamic gesture recognition both using the combination of LSTM networks and the data collection framework based on cameras. It captures gesture via camera, extracts key points of the hand using MediaPipe, and trains an LSTM model to classify gestures according to predefined actions. Therefore, it increases accuracy as well as allows for interaction to be applied in the communication through gesture.

Keywords: Real-time gesture recognition · CNN-LSTM hybrid model · dynamic hand gestures · human-computer interaction (HCI) · sign language recognition

1 Introduction

Though rapidly increasing advances of machine learning and computer vision, there is yet significant challenges in the direction of real-time dynamic gesture recognition. Most of the currently available approaches would depend on either CNN-based image-based gesture detection or RNNs to be able to account for time information. Applying these models to real-time recognition tasks would result in several trade-offs between latency and accuracy.

More recent studies have demonstrated that CNNs have been very effective for fixed images in classification tasks. However, they don't learn the movement flow with time in dynamic sequences. Hence, this work combines an LSTM network with feature extraction via MediaPipe to capture hand detailed landmarks and body posture; we attempt to improve upon the model's capacity in distinguishing many gestures through an improved temporal accuracy in the work.

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 38–53, 2026.
https://doi.org/10.1007/978-3-031-98138-8_4

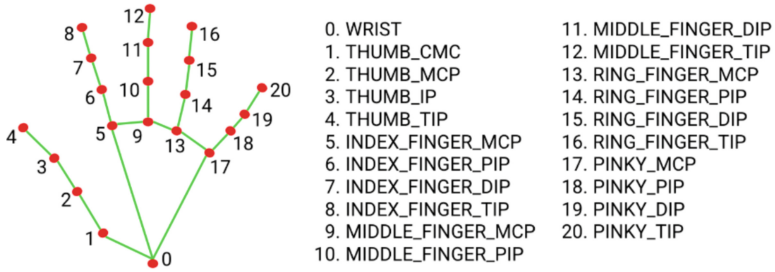


Fig. 1. Mediapipe Hand Detection Key points

The application of our model includes real-time support for sign language recognition, which can be revolutionary in accessibility in different contexts. The study focuses on a dataset of dynamic gestures and explores how sequential learning models can capture the subtlety of hand movements more effectively than traditional models. The hand movements are tracked with the media-pipe hand detection module for which key relative co-ordinates are tracked as shown in Fig. 1. We aim to deliver a fast and accurate gesture recognition solution suitable for real-time deployment through a combination of data pre-processing, efficient feature extraction, and sequence learning.

2 Literature Review

The application of Convolutional Neural Networks (CNNs) has significantly enhanced the recognition of static gestures [1]. Methods like colour segmentation and data augmentation have demonstrated high accuracy, even under challenging background conditions, making them suitable for tasks requiring stationary gesture recognition [2]. However, CNNs struggle with dynamic gestures as they fail to capture temporal patterns, which are crucial for recognizing movements across time [4]. To address this, Long Short-Term Memory (LSTM) networks are integrated due to their ability to capture sequential dependencies, especially in dynamic gestures, such as those used in sign language [4, 5]. Hybrid models that combine CNN and LSTM layers have been developed to address the need for both spatial and temporal feature extraction. These models employ CNNs for spatial data and LSTMs for sequence processing, proving effective in recognizing complex gestures. However, they present increased computational demands, posing challenges for real-time applications [5, 6]. Further advancements have been made with 3D-CNNs and LSTM networks, reducing computational loads while effectively handling spatial and temporal features, enabling real-time functionality on intricate gesture datasets [6, 7].

MediaPipe's holistic tracking system is widely used in gesture recognition as it offers accurate key point extraction for hand and body landmarks. By delivering precise, low-latency tracking, MediaPipe enhances machine learning models, making it ideal for real-time gesture recognition tasks [1, 7]. When combined with LSTM networks, MediaPipe further strengthens model robustness, particularly in applications that require accurate hand movement recognition [7, 8]. Gesture recognition holds significant potential for accessibility, such as real-time gesture interpretation for hearing-impaired individuals

in both physical and virtual spaces, requiring high accuracy across diverse backgrounds and lighting conditions. This highlights the importance of creating user-independent systems that perform consistently across various scenarios [1, 8].

Recent studies support the use of LSTM networks for sequential data in gesture recognition. For example, Kim et al. [9] proposed a dynamic hand gesture recognition model using MediaPipe, Inception-v3, and LSTM, showing improvements in gesture recognition accuracy. Similarly, Kumar et al. [10] explored the combination of MediaPipe and CNNs for real-time American Sign Language (ASL) gesture recognition. Lee and Kim [11] presented LM-Net, a dynamic gesture recognition network that captures long-term temporal context, enhancing performance in complex gestures. Studies by Chen et al. [12] and Kumar et al. [13] further illustrate the benefits of CNN-LSTM integration, particularly for tasks like drone control and dynamic hand gestures, where both spatial and temporal patterns are essential.

Other advancements include hybrid models like CNN-BiLSTM, which have been used for sign language recognition with higher accuracy [15, 21]. Further, researchers like Singh and Yadav [24] and Ahn and Han [25] investigated LSTM-based models for real-time applications, such as augmented reality (AR) and multimodal gesture recognition, demonstrating LSTM's efficacy in dynamic gesture prediction and multimodal environments. In conclusion, literature shows that combining MediaPipe for landmark extraction with LSTM networks is promising for real-time gesture recognition. Although CNN and hybrid models provide benefits, they face limitations with sequential data, making LSTM an essential component for achieving high accuracy in dynamic gesture recognition systems.

3 Theoretical Framework

Dynamic gesture recognition involves a more complex setup such as real-time video capture, high-precision key-point extraction, and sequential data processing. This is aimed at handling complex gestures through the inference of slight movements and positions of body parts such as hands and upper bodies. The core backbone of this approach leans on MediaPipe's holistic tracking model and Long Short-Term Memory (LSTM) neural networks. Feature is extracted from the dataset and passed into the trained model which on receiving processed real time input from webcam generates a gesture output based on probability of prediction Fig. 2. Each of these modules works distinctly in developing a system that correctly perceives gestures in realistic environments with any background or illumination and noise.

3.1 Mediapipe Holistic Model

First and foremost, the holistic tracking system that MediaPipe gives is a very important factor in this framework. It is an advanced computer vision model that can map landmarks and detect them across several body parts from hands to face and pose.

The hands are the most fundamental gestures in gesture recognition because the communication weight is carried heavily through them. MediaPipe's holistic model is well able to identify hand and pose landmarks with a much better accuracy and

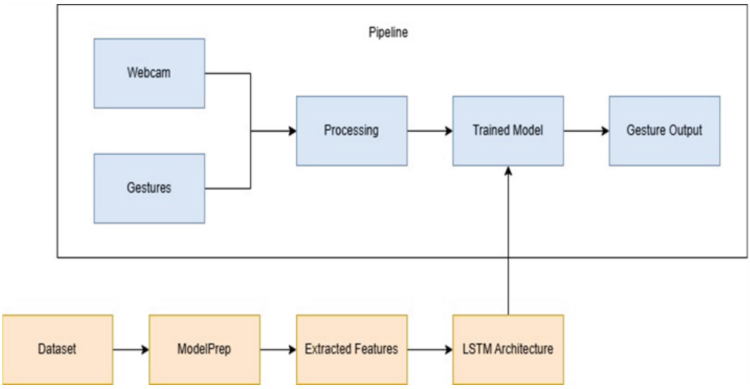


Fig. 2. Flow of Data

also provides the benefits of real-time tracking that has minimal latency, but in gesture dynamics, changing speed of hand position orientation comes into play.

3.2 Key Point Extraction

The holistic model will be adapted such that the hand and pose landmarks mainly form the basis on which extraction of critical points might be focused. Consistent key points are produced by the holistic model for both left hand and right hand and thus give the system a capability to handle symmetrical or mirrored gestures, characteristic of gestures in sign languages and communication. The extraction of key points is the core of the gesture recognition system, translating visual data into a numerical format for machine learning. Key points are specific positions on the hands and pose, tracked over a sequence of frames that captures the spatial configuration of the body as it moves.

3.3 LSTM Neural Network

For every hand and body movement, gestures have a definite number of frames through which these actions are coordinated. Frames assembled together are further used in the process for recovery with 3D coordinates, including x, y, and z positions of landmarks found. These are taken and filled into arrays that are readily fed into the model during gesture recognition. After key points are extracted, the system then processes the sequence of landmarks over time using an LSTM neural network. LSTMs are unique from other networks because they can remember previous inputs; this makes them ideal for data that is temporal in nature, such as gesture sequences. The LSTM can remember previous frames. Therefore, the model will be able to recognize flow and rhythm in gestures like a slow lift of the hand or a slight bend of fingers. This allows the system to distinguish between gestures that look similar but have different motions over time. With LSTMs, the model can reach a very good accuracy in gesture classification, even in complex scenarios.

3.4 Real-Time Prediction with Thresholding

To have a smoother flow, the system involves a real-time prediction mechanism based on a threshold-based approach. It utilizes the combination of a cooldown period along with skip-frame techniques in stabilizing the predictions where the system will not reproduce similar or wrong results. A cooldown counter for consistency in output will prevent further predictions when a gesture is detected. High confidence, predictions greater than the set probability threshold, increase the robustness of the system while avoiding false positives. In live settings such as a public space or interactive virtual environments, real-time settings can ensure a proper run.

4 Implementation

4.1 Static Signs

Key Point Extraction: The static gesture recognition system uses MediaPipe's key point extraction. This is a machine learning library optimized for real-time landmark detection. The core of the extraction lies within the preprocessing.py script, which actually uses the Hands model from MediaPipe to detect hand landmarks or key points. Every hand landmark is mapped to the corresponding x, y coordinates that are spatial positions in each image frame.

Image and Landmark Processing: The function calc_landmark_list calculates the positions of landmarks for each hand on an image from raw landmark data from MediaPipe into coordinates in pixels. For each landmark, the relative position is necessary for the interpretation of the hand gesture and capture of hand orientation, finger position, and shape variations.

Normalization: The pre_process_landmark function normalizes all landmark's coordinates to be relative to a base point, commonly the wrist; this eliminates the influence of size and orientation on hands. Another thing, the landmarks were flattened to a one-dimensional array for the efficient input in the neural network to further spatial pattern recognition.

Data Augmentation: This will improve generalization since the function augment image horizontally flips the image, mirroring gestures to create more samples for training.

Data Collection: Similarly, key point is extracted while saving to CSV for training in preprocessing.py and each processed image, the key points are labelled with their id corresponding to the certain static gesture that is expressed using either letters or numbers.

Gesture Labels: Each move obtains a name-for example, "A" or "1"-designating the class for further classification. Names correspond to class names used during the training period.

Data Logging: Key points with corresponding gesture labels are stored in structured form in keypoint.csv Fig. 3, resulting in an orderly dataset on which the models will train. This system facilitates ready access and systematic handling of samples for training.

Dataset: To collect images for the static gesture recognition system, the target scope aims to capture all significant points of hand positions depicting any letter from A-Z or numbers from 0 to 9. This will thus ensure that diversity of such gestures ensures generalizability with cross-use, as in different users and possibly different conditions.

Gesture data was captured using a high-definition camera set to a resolution of 1920 $\times$ 1080 pixels. To ensure consistency and reproducibility, the data collection process was conducted in a controlled environment with uniform lighting conditions to minimize shadows and reflections. A neutral background was used to reduce noise and enhance keypoint detection accuracy. MediaPipe was employed to extract hand and pose keypoints from each video frame. Each gesture sequence comprised 20 frames, representing dynamic movements over time.

Data Collection and Structure: There are about 1,000 images per gesture class. Thus, the total is 36,000 images. Each image is a unique gesture corresponding to a specific letter from A to Z or a number from 0 to 9. Thus, this well-balanced dataset ensures having enough samples for each class to train effectively in a multi-class classification model.

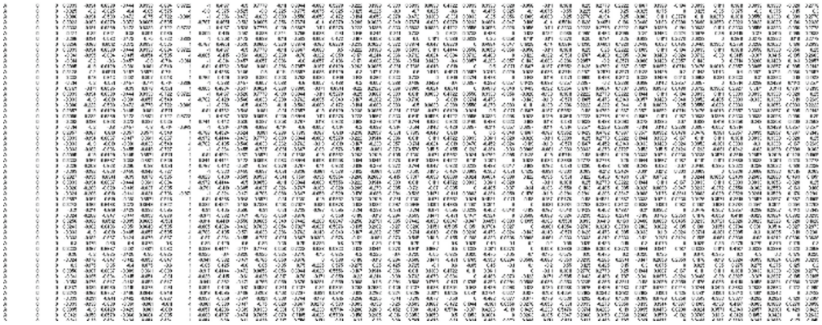


Fig. 3. Head CSV file with logs

Image Preparation and Pre-processing: Every picture is capturing hand in one specific static pose. All the images were captured with rather different illumination conditions as well as changes in hand position for adding augmentations.

The images were pre-processed as follows:

- Key point Extraction: The MediaPipe Hands model extracts the hand landmarks from an image and encodes them as structured arrays of x, y coordinates for each key point.
- Normalization Key points are normalized relative to a base point (wrist position). Thus, the model is not sensitive to slight hand size and orientation movements.
- Data Augmentation: To increase generalization, all images were flipped horizontally, and the effective size of the dataset was doubled to 2,000 samples per class. This augmentation ensures that the model identifies both left and right hands of gestures.

Data Organization: The dataset has directories labelled with class labels for each unique gesture. There would be efficient loading and retrieval of data due to this structure of the directories.

- Root Directory: The main directory (data/) holds all gesture images.
- Subdirectories: Each gesture class (e.g., A, B, 1, 2) has its own subfolder within the main directory. Within each class folder, individual image files are named sequentially (e.g., data/A/0.jpg, data/A/1.jpg, etc.) for streamlined access.
- CSV Logging: Key points and labels are stored in keypoint.csv, each row per image's key points with a corresponding gesture label. This structure allows easy batching and model input during training.

Consistency and Quality Control:

- Root Directory: The main directory (data/) holds all gesture images.
- Subdirectories: Each gesture class (e.g., A, B, 1, 2) has its own subfolder within the main directory. Within each class folder, individual image files are named sequentially (e.g., data/A/0.jpg, data/A/1.jpg, etc.) for streamlined access.
- CSV Logging: Key points and labels are stored in keypoint.csv, each row per image's key points with a corresponding gesture label. This structure allows easy batching and model input during training.

Model Architecture: A densely connected neural network model is designed to classify static gestures based on hand landmarks. The architecture was implemented using TensorFlow and Keras, structured to recognize spatial patterns within hand landmarks.

Model Layers:

- Input Layer: It accepts pre-processed arrays of hand landmarks as input.
- Hidden layers: There are multiple fully connected dense layers that learn the high-level spatial patterns over static hand positions. It has dropout layers to regularize and reduce overfitting.
- Output Layer: Softmax -activated dense layer produces class probabilities for each input corresponding to some gesture.

Compilation:

- Input Layer: It accepts pre-processed arrays of hand landmarks as input.
- Hidden layers: There are multiple fully connected dense layers that learn the high-level spatial patterns over static hand positions. It has dropout layers to regularize and reduce overfitting.
- Output Layer: Softmax -activated dense layer produces class probabilities for each input corresponding to some gesture.

Training Process: The model is further trained with their extracted hand landmarks and corresponding labels in a supervised manner. Data is then split into training and testing sets to enable generalizable learning.

Training Parameters: Training: It trains on 50 epochs with a batch size of 128, and tracks its performance on an 80–20 validation split.

Early Stopping: Training with validation: Early stopping based on validation loss Prevents overfitting, by halting training when no further improvements have been detected.

Model Evaluation: The performance of the model is checked on the test set on how well it identifies the unseen static gestures. Performance metrics include accuracy, precision, recall, and F1-score.

Prediction and Classification: The model evaluates class probabilities for each test sample and predicts classes based on the probability distribution over gesture classes. It then compares classes against the corresponding ground truth labels to compute evaluation metrics.

Metric Calculation: Standard evaluation metrics are accuracy, precision, recall, and F1-score providing a full view of the model capability in classifying stationary gestures.

Real-Time Prediction: The script realtime.py loads the trained model for real-time gesture recognition through a webcam interface. This model essentially allows people to see the predicted gestures overlaid onto the video feed.

Gesture Detection Pipeline:

- Key Point Extraction. During processing of the frame, MediaPipe captures key points.
- Normalization: Key points are normalized just like in the training setup for fair recognition.

Gesture Prediction and Visualization: During computation, the model infers gesture in real time; with recognized gestures appearing onscreen to visually provide an immediate feed-through loop for live user interactivity.

4.2 Dynamic Gesture

Key Point Extraction: KeypointsExtraction.py The main script involved in transforming the raw video frames into meaningful data is that of MediaPipe's holistic model, which can detect hand and pose landmarks in real time. It maps the detected landmarks onto x, y, and z coordinates within each frame, thereby representing the spatial position of each key point on the hands and upper body.

Key point extraction would hold two principal functions which are image process and keypoint_extraction _with position. The first one, the image process would prepare every frame for processing by MediaPipe, while the latter, keypoint_extraction_with_position would draw out the exact location of each landmark. Therefore, this would represent the relative position of hand within the frame that would give a spatial context. For example, a raised hand, bent finger, or other type of hand gesture would be possible by the specific patterns in these coordinates. It converts all these coordinates into a single array for effective storage of spatial data about every gesture sequence.

Data Collection: DataCollection.py will be the most central part of collecting gesture data for model training. Within this, sequences of frame captures through a webcam are being taken. The process is designed to create labelled datasets for each gesture that

trains the gesture recognition model. MediaPipe Holistic has been added to the system to capture all major key and significant hand and body landmarks in the process in order to track motion and positioning better for gestures.

4.3 Setup and Initialization

Gesture Labels: A list of gesture labels is defined, which are later used to denote class names for model classification.

Directory Structure: The PATH variable defines the root folder for the storage of data. For every action of gestures, it creates a subfolder that stores all the recorded sequences. For example, frames of each sequence are stored in distinguished directories in the folder structure data/Hello/0, data/Hello/1, etc., so that they can be accessed easily at training time during analysis of data.

Camera Initialization: The code checks if the camera is accessible (cap.isOpened()). If not, an error message appears, terminating the script gracefully. It verifies whether harvesting data can be done without any interruption considering camera unavailability.

Data Collection Process: Each gesture is captured in a controlled sequence of frames (20 frames per sequence in this configuration, with every frame recorded and labeled as a unique part of the gesture sequence).

Recording Control: When the script asks the user to begin recording a particular gesture, the space bar is pressed to initiate the recording of the gesture and sequence. This setup allows manual control; it gives user time to prepare the gesture and ensures only intended frames are being captured, which reduces noise in datasets.

Frame Capture and Processing: For every frame, MediaPipe Holistic uses image_process(image, holistic) for processing a webcam image so that results with landmarks are given. The script overlays landmark points on the camera feed using draw_landmarks(image, results), which gives the user visual feedback to achieve better accuracy.

Key point Extraction with Position: For each frame, keypoint_extraction_with_ position(results, image.shape) retrieves the landmarks with coordinates x, y, and z based on the size of the image. The spatial context in these movements **is greatly important to the model for it to get the spatial structure of gestures.**

Frame Saving: Each frame's key points are saved as a.npy file, capturing the landmarks for each sequence of the gesture. It is nimble and easy to handle since separate frames keep it separated and certain frames can be recovered or left out without altering other data.

Error Handling and Exit Options: It has 'q' to quit recording to abort the program at any moment from the recording making the data capture incomplete. This flexibility allows the user to abort and restart specific gestures if errors occur, promoting a clean dataset. This just leads to an ordered, homogeneous dataset of gesture sequences of equal frame counts, with the model able to train nicely and get truly effective, well-labeled inputs.

The Dataset: The resultant dataset will be a well-organized collection of labeled gesture sequences, structured to facilitate effective model training. Each gesture is recorded in sequences of frames, with each sequence stored in its respective subdirectory. The dataset is designed with the following attributes:

Structure and Storage: Root Directory: The main folder (data/) contains all gesture data, with each gesture type organized under subfolders. Subdirectories: Each gesture (e.g., "Hello") has its own folder, further divided into sequential subdirectories (data/Hello/0, data/Hello/1, etc.) where each subdirectory represents an individual gesture instance.

Frame Data: Each sequence is made up of 20 frames, each stored separately to maintain consistency across all gesture sequences.

Data Consistency: Homogeneity: Each gesture has an equal frame count, creating a uniform dataset that is easier to batch and feed into the model.

Coordinate Arrays: Each frame's landmarks are stored as.npy files, containing a structured array of x, y, and z coordinates. This format enables quick loading and processing without reformatting.

Labeling and Classification: Gesture Labels: Every gesture sequence is labeled, providing clear class names for each gesture type, which serves as the target variable for the model.

Spatial Context: The coordinates in each frame capture the relative positioning of key points, offering a spatial context that the model can leverage to distinguish gestures based on body and hand movements.

Error Handling and Flexibility: Manual Recording Control: Users can initiate and end recordings, reducing unwanted noise in the dataset by ensuring only intentional gestures are recorded.

Error Tolerance: Any interrupted or erroneous recordings can be restarted, supporting a clean, high-quality dataset.

Each array captures the spatial position of landmarks, giving a snapshot of the gesture in that frame Fig. 4. When saved as.npy, this array can be easily loaded in Python. This organized and labeled dataset of gesture sequences, with consistent spatial context, will enable the model to learn and distinguish various dynamic gestures effectively. The captured key point arrays provide rich data for both training and validation, setting a solid foundation for dynamic gesture recognition.

Model Training: The ModelTraining.py script develops a CNN-LSTM hybrid model, which is supposed to extract both spatial features, that is, from hand landmark coordinates, and temporal patterns across frames, which are essential for accurate gesture classification. Such an architecture is beneficially suited for dynamic gestures, as it captures local spatial dependencies while learning sequential patterns. Here's a detailed explanation of each layer and what they do in the model to handle gestures represented by hand landmarks.

Model Architecture: The architecture uses a Sequential model from Keras as follows Fig. 5.

```
[
    [0.45, 0.77, -0.05],   # Right wrist
    [0.48, 0.72, -0.04],   # Right thumb CMC
    [0.50, 0.68, -0.03],   # Right thumb MCP
    [0.53, 0.65, -0.04],   # Right thumb IP
    [0.55, 0.63, -0.03],   # Right thumb tip
    ...
    [0.60, 0.85, -0.01],   # Left shoulder
    [0.58, 0.90, 0.0],     # Left elbow
    ...
]
```

Fig. 4. A sample npy file storing relative coordinates with the frame

Input Layer: Shape: (frames, 126), representing the number of frames that make up each gesture sequence, where 126 depicts the flattened 21 key points for every hand, x, y, and z coordinates, respectively.

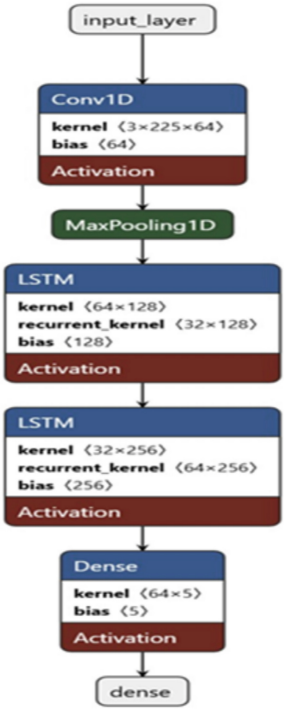


Fig. 5. Model Architecture

Conv1D Layer: The given Conv1D layer of 64 filters with kernel size 3 extracts spatial patterns from each frame of the input hand landmarks. Activation: relu, to introduce non-linearity so the model learns intricate spatial features across all key points.

MaxPooling1D Layer: This layer reduces the spatial dimension, thus enabling the model to focus on key spatial features while at the same time minimizing computational load. Pooling Size: 2, which will reduce the dimensionality and help in generalization.

First LSTM Layer: Units: 32, return_sequences = True so that the next LSTM layer receives the entire sequence of processed features. Activation: relu, which enables the LSTM to keep information regarding sequential dependencies of the gesture.

Second LSTM Layer: Units: 64. Final Summary of the entire sequence. Outputs a distilled representation that captures the time evolution of the gesture.

Dense Output Layer: Purpose of softmax-activated dense layer is the classification of gesture into one of the pre-defined gesture classes.

Neurons: The number of neurons maps directly to the number of gestures classes and provides a probability distribution for classification.

Model Compilation: The model is compiled with-

- Optimizer: Adam, and it efficiently manages a dynamic learning rate for smooth convergence.
- Loss Function: categorical_crossentropy: for multi-class classification.
- Metrics: accuracy, allows the model to keep track of its performance in classifying gestures during training.

Training Process: The dataset, containing hand landmark sequences, is now loaded and split into training and testing sets. The model will go through 100 epochs so that it could refine its parameters toward better accuracy in recognizing gesture sequences.

Model Evaluation: Then, it evaluates the trained model on the test set to know about the accuracy of the model in real-world scenarios. Predictions are generated, and accuracy is calculated by comparing the predicted labels with actual test labels.

Real Time Prediction: RealTime.py module is dealing with actual prediction of gesture in a real-time way by getting loaded the model that has learned with all kinds of poses along with opening the Webcam in the screen. Minimize the prediction stability also there with cool down mechanisms with a few frames ahead from the frame where key points is captured as the whole 20 frames. Output comes only at highly confident key points through gestures prediction. These enhancements have significantly improved Realtime accuracy.

5 Results

The static and dynamic gesture recognition models demonstrated strong performance across various metrics, validating the data processing pipeline, dataset quality, and model architecture. A user-independent system is crucial for real-world applications, as it

ensures that the model performs reliably regardless of individual user differences in hand shape, size, and gesture execution style.

5.1 Static Signs Detection

For Static Signs, the alphabet and number gestures trained on the model had high accuracy and performance consistency across the key metrics. It reached an accuracy of 0.984, with a precision of 0.987, recall of 0.986, and an F1-score of 0.986. These metrics prove how the model may classify a wide variety of static gestures in different hand poses and lighting conditions. A higher value in precision would indicate a low rate of false positives wherein the model correctly identifies the gesture but mistakenly classifies other inputs as gestures. The same high recall score validated the model's reliability towards correctly detecting the gestures with fewer instances of missed gestures. The balance in precision and recall as reflected through the F1-score necessarily implies that the model is robust and suitable for a real-world static gesture recognition system. Table compares the performance of our model with the other existing models (Table 1).

Table 1. Performance comparison of the proposed model.

Parameter/ Algorithm	Proposed Approach	Eid, A., & Schwenker, F. (2023)	Prakash, A. J., et al. (2022)	Rehman, M. U., et al. (2022)
Accuracy	0.984	0.982	0.956	0.964
Precision	0.987	0.967	0.948	0.950
Recall	0.986	0.975	0.932	0.945
F1-score	0.986	0.971	0.940	0.947

5.2 Dynamic Signs Detection

The LSTM model showed perfect accuracy during training, reflected in the classification report as performance over the dataset. This is based on predictions of training data rather than seeing the data, thus reflecting that the model has learned from the data effectively.

During real-time testing, it performed very well with the efficiency of gesture prediction without a delay Fig. 6. The cooldown mechanism and hand-detection check were incorporated into the design of the prediction logic such that unwanted or duplicate predictions were avoided. Through this external logic, it offered smooth, accurate outputs and was not interrupted by quick hand movements or changes in positions, thus enhancing user experience.

These results were achieved with the help of data collection using MediaPipe Fig. 7, which was able to pick up precise hand key points. This allowed for a comprehensive and reliable dataset, which helped the model generalize well to real-time scenarios and ensure that it could handle dynamic hand gestures with accuracy and speed.

Fig. 6. Real Time Sentence Formation

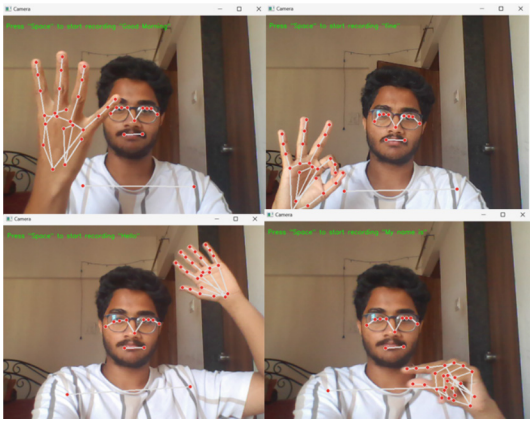


Fig. 7. Data collection process

Overall, the results validate the system's ability to recognize both static and dynamic gestures effectively. The static gesture model's high-performance metrics confirm the reliability of the hand landmark data and neural network architecture, while the dynamic model's real-time adaptability highlights the effectiveness of the LSTM design and prediction logic. Together, these components establish a solid foundation for gesture-based communication, achieving robust and accurate recognition across a wide range of gestures.

6 Conclusion and Future Scope

This project on dynamic gesture recognition holds great promise for applications across domains. One of the promising areas for future work is in improving accessibility technology, especially real-time translation for sign language. Good groundwork has been

laid down by the current models; however, there is room to expand capabilities in recognizing more complex gestures and even finger-spelling for full vocabulary coverage in sign languages. Future implementations could be considering the use of this model as a basis for recognizing multilingual sign languages because it could be adapted to apply to the different nuances from one region to another.

More advanced sensor data would be incorporated, like depth information from sensors such as the Microsoft Kinect or Apple's LiDAR, which could offer even richer spatial data than 2D images. This would enhance the accuracy of gestures with complex depth cues and increase robustness against varying lighting and background conditions. Depth-based data can let the model operate with richer three-dimensional gestures, thus becoming more suitable for the applications of virtual reality and augmented reality, which are often mostly gesture-driven.

References

1. Eid, A., Schwenker, F.: Visual static hand gesture recognition using convolutional neural network. Algorithms **16**(8), 361 (2023)
2. Prakash, A.J., Pławiak, P., Samantray, S.: Real-time hand gesture recognition using fine-tuned convolutional neural network. Sensors **22**(3), 706 (2022)
3. Pławiak, P., Samantray, S.: Real-time dynamic gesture recognition in human-computer interaction. IEEE Trans. Syst. Man Cybern. Syst. (2021)
4. Rehman, M.U., et al.: Dynamic hand gesture recognition using 3D-CNN and LSTM networks. Comput. Mater. Continua **70**(3), 4675–4690 (2022)
5. Zheng, J., et al.: Hybrid CNN-LSTM model for dynamic gesture recognition. IEEE Trans. Multimedia (2020)
6. MediaPipe. MediaPipe holistic tracking system for real-time applications (2022)
7. Pathak, R., et al.: Gesture recognition in accessibility: enhancing communication for the hearing-impaired. J. Hum.-Comput. Interact. (2023)
8. Zhang, Y., Li, M.: Dynamic gesture recognition using stacked LSTM networks. IEEE Xplore (2021)
9. Kim, J., Jamil, S., Lee, J., Ullah, F.: Next-Gen dynamic hand gesture recognition: MediaPipe, inception-v3, and LSTM-based enhanced deep learning model. Electronics **13**(16), 3233 (2024)
10. Kumar, R., Bajpai, A., Sinha, A.: Mediapipe and CNNs for real-time ASL gesture recognition. arXiv preprint (2023)
11. Lee, H., Kim, J.: LM-Net: a dynamic gesture recognition network with long-term temporal context. SpringerLink (2021)
12. Chen, L., et al.: Real-time hand gesture recognition using CNN and LSTM for drone control applications. MDPI Sens. **22**(9), 4598 (2022)
13. Kumar, N., Kumar, S., Kumar, M.: Dynamic hand gesture recognition using CNN. SSRN Electron. J. (2023)
14. Jaeho, K., et al.: Facial expression recognition using hybrid CNN and ConvLSTM networks. SpringerLink (2022)
15. Kumar, A., Bajaj, P., Verma, R.: Enhancing sign language recognition: a CNN-BiLSTM approach. IEEE Xplore (2023)
16. Javed, F., et al.: Assessing the influence of LSTM and post-processing in CNN-based gesture recognition. SSRN Electron. J. (2023)

17. Rehman, S.: Enhancing the accuracy of gesture recognition using CNN. IEEE Trans. Hum.-Mach. Syst. (2023)
18. Liu, Z., Wang, H.: Gesture recognition in human-computer interaction: challenges and solutions. IEEE Trans. Syst. Man Cybern.: Syst. (2023)
19. Wang, F., et al.: Comparative analysis of CNN and LSTM for continuous hand gesture recognition. IEEE Trans. Hum.-Mach. Syst. (2023)
20. Park, J., Kim, Y.: High-efficiency real-time gesture recognition using CNN-LSTM networks. Springer Nature (2024)
21. Chen, X., Zhang, S.: A CNN-BiLSTM based multimodal continuous hand gesture recognition system. PLOS ONE (2023)
22. Ahmad, S., et al.: Data glove-based gesture recognition using CNN-BiLSTM model. PLOS ONE (2022)
23. Lin, D., Chen, Z.: Real-time dynamic gesture recognition: combining CNN and LSTM for improved accuracy. SpringerLink (2023)
24. Singh, P., Yadav, A.: Using LSTM networks for real-time gesture prediction in AR applications. IEEE Xplore (2023)
25. Ahn, Y., Han, J.: Real-time multimodal gesture recognition using CNN and LSTM integration. MDPI Sens. **22**(6), 1105 (2022)

SpeechCraft: Modular AI Conversation System Using Multivariate LLMs

Vishnu Kamisetti(✉), Goutham Bittla, Dhanunjai Gujjaboina, Katta Sugamya, and Thimmapuram Madhuri

Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, India

vishnu.kams@gmail.com, {ksugamya.it,tmadhuri.it}@cbit.ac.in

Abstract. SpeechCraft constitutes an innovative, modular approach towards the development of conversational AI systems. It provides a flexible framework for developing customizable voice assistants and experimenting with state-of-the-art models in speech recognition, natural language processing, and text-to-speech synthesis. The architecture is divided into three major components of the system: speech-to-text conversion, text processing for response generation using a large language model, and text-to-speech synthesis. This can be independently configured in such a way that users can choose from a range of options, including popular APIs such as OpenAI, Groq, Deepgram, and local model implementations. The key features are easy audio recording and playback, central configuration management for ease, quick prototyping, the ability to compare different models of AI, and adaptations to varied use cases; further it extends its adaptability towards language-specific models for correct processing and regional language and dialect generation. Performance evaluations reveal that the system achieves an average latency of 2.5 s, end-to-end task accuracy of 93%, and a user satisfaction score of 4.6/5, making it adaptable to diverse language and regional requirements. SpeechCraft is a valuable tool for researchers and developers to advance conversational AI systems.

Keywords: Conversational AI · Speech Recognition · Text-to-Speech Synthesis · Configuration Management

1 Introduction

The SpeechCraft project is an innovative modular AI system designed to streamline the development of customizable voice assistants. It aims to address challenges in conversational AI by offering a flexible, easy-to-configure framework that facilitates experimentation with cutting-edge models in speech recognition, natural language processing (NLP), and text-to-speech (TTS) synthesis. The architecture of the system consists of three core components: speech-to-text, text processing, and TTS synthesis. These components can be configured independently, allowing users to select from a variety of APIs and models such as OpenAI, Groq, and Deepgram. By supporting rapid prototyping

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 54–62, 2026.
https://doi.org/10.1007/978-3-031-98138-8_5

and comparative analysis, SpeechCraft empowers researchers and enthusiasts to create adaptable AI systems for various language and regional needs. The project ultimately improves accessibility to advanced AI for broader application and innovation.

2 Literature Review

2.1 Related Work

1. Malodia et al. [1] examined the impact of AI voice assistants such as Siri and Alexa on consumer behavior, finding that trust and personalization significantly influence user satisfaction. However, their study lacks information on regional adaptability and modularity.
2. Kraus et al. [2] highlighted the role of proactive conversational assistants in education, emphasizing their potential to improve learning outcomes. Their work identifies privacy concerns but does not address modular frameworks.
3. Pal et al. [3] explored emotional attachment to voice assistants through the lens of Sternberg's Triangular Theory of Love. While they provide insights into user engagement, the study overlooks the technical implementation of modular systems.
4. Babu et al. [4] detailed advancements in transformer-based NLP architectures, showcasing their applications in conversational AI. However, scalability and computational costs remain challenges.
5. Gowthamy et al. [5] presented innovations in TTS synthesis using machine learning, focusing on enhancing user experience. Yet, their study does not consider customization for regional languages.
6. Joshi et al. [6] developed a personal desktop assistant using voice recognition and NLP, demonstrating improved productivity. The project's narrow focus on desktop applications limits its applicability to broader conversational AI contexts.
7. Baker et al. [7] introduced the Intelligent Voice Instructor-Assistant System (IVIAS), which supports large classroom environments. While their system addresses scalability, it lacks flexibility for diverse use cases.
8. Pal et al. [8] investigated privacy concerns in voice assistants, using a privacy calculus framework. Their findings underscore the trade-offs between personalization and privacy but do not explore modular solutions to address these challenges.
9. Dutsinma et al. [9] systematically reviewed the usability of voice assistants, advocating for standardized evaluation metrics. Their study's focus on usability metrics provides a foundation for further exploration in modular frameworks.
10. Seaborn et al. [10] surveyed advancements in voice interaction for humanagent communication, identifying gaps in conversational dynamics. Their recommendations align with the need for adaptable systems like SpeechCraft.

2.2 Research Gap

Existing systems often lack modularity and adaptability for regional languages and user-specific needs. SpeechCraft addresses this gap by providing a scalable framework for integrating diverse AI models and APIs, promoting experimentation and customization.

Table 1. Summary Table of Studies

Study	Key Outcomes	Limitations
[1]	Enhanced user trust and engagement	Limited regional adaptability
[2]	Improved learning outcomes through AI	Privacy concerns
[3]	Emotional attachment to VAs	Overreliance on specific datasets
[4]	Transformer-based NLP advancements	High computational cost
[5]	TTS synthesis innovations	Scalability issues
[6]	Productivity improvement through desktop assistants	Limited application scope
[7]	Scalability in classroom settings	Lack of flexibility for diverse use cases
[8]	Insightsintoprivacy-personalization trade-offs	Absence of modular solutions
[9]	Standardized usability metrics for VAs	Limited focus on modular frameworks
[10]	Advancements in voice interaction dynamics	Gaps in conversational adaptability

2.3 Summary Table

(See Table 1)

3 Objectives

The objectives of SpeechCraft are to provide an agile and modular framework for building customizable voice assistants, simplify integrations with a wide range of advanced AI models for different dialect and language needs, enable rapid prototyping and comparison of various conversational AI configurations. This design enables to personalize user-voice assistant most efficiently in the context of their wide use cases.Key Objectives of the system are:

3.1 Develop a Modular Framework

Create a flexible system for building customizable voice assistants, allowing easy configuration of components like speech recognition, NLP, and TTS.

3.2 Integrate Advanced AI Models and APIs

Seamlessly integrate leading AI models and APIs to support speech to text, language processing, and text-to-speech(tts) functionalities for diverse use cases.

3.3 Enable Comparative Analysis and Prototyping

Provide tools for real-time comparison of different AI models, enabling rapid prototyping and performance evaluation.

3.4 Support Language-Specific Models

Implement language-specific models to improve the accessibility and accuracy of voice assistants for regional languages and dialects

4 Methodology

4.1 System Architecture

By dividing the architecture into separable but integrable components, such as speech-to-text conversion, response generation, and text-to-speech synthesis, so the stage of conversation can be easily configured and optimized. This design approach enables developers to adapt the system based on specific project needs, facilitating experimentation with a range of models and APIs for enhanced versatility.

1. **Speech-to-Text Conversion**: Handles the process of converting spoken audio into text. Users can choose from various speech recognition models and APIs, such as Google Speech-to-Text, AWS Transcribe, or on-device models like DeepSpeech.
2. **Response Generation**: The natural language processing (NLP) module is responsible for understanding the user's input and generating an appropriate response. SpeechCraft supports the integration of large language models, such as GPT-3 or BERT, as well as custom-trained models.
3. **Text-to-Speech Synthesis**: The final component converts the generated response text back into spoken audio, enabling the voice assistant to communicate with the user. It provides support for multiple text-to-speech engines, including cloud-based services like Amazon Polly, Google Cloud Text-to-Speech, and local TTS models.
4. **Data Flow Between Modules**: Each module is interconnected via an asynchronous message queue, ensuring seamless data transfer and parallel processing. Error logs and status updates are maintained in a centralized dashboard.
5. **Error Handling**: If the STT module fails to produce accurate transcriptions, the system requests re-input or switches to a backup model. Similarly, fallback strategies are in place for NLP and TTS modules.

The system supports both online and offline configurations, making it adaptable to various environments (Fig. 1).

4.2 Implementation Details

The system is implemented using Python, with APIs for speech processing and NLP. Key configurations include:

- **Hardware**: Intel i7-12700H CPU, NVIDIA RTX 3080 GPU, 32GB RAM.
- **Software**: Python-based implementation with libraries such as PyTorch, TensorFlow, and librosa.
- **Datasets**: LibriSpeech for STT, VCTK for TTS, and OpenSubtitles for NLP tasks.

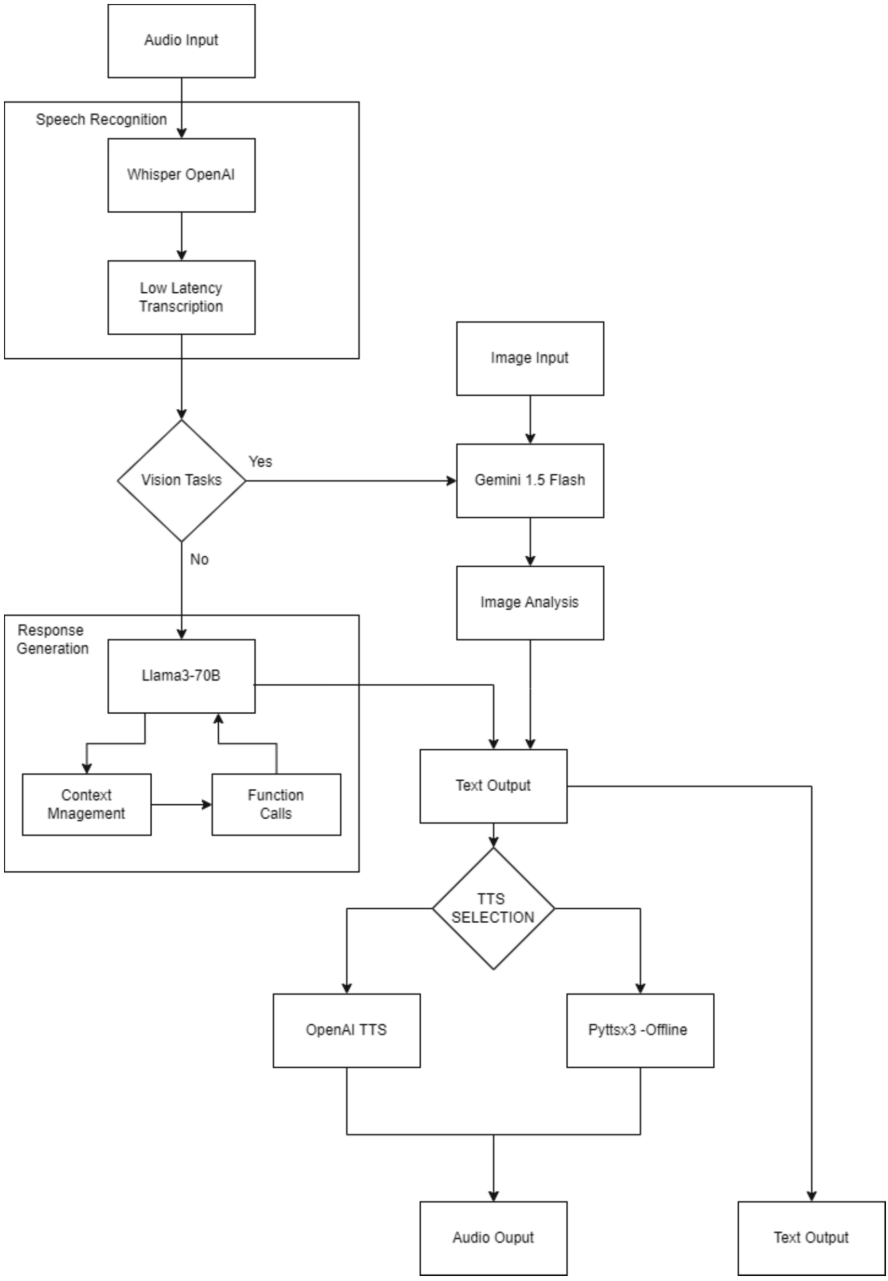


Fig. 1. System Flow Diagram of the Proposed Model

5 Evaluation Metrics

The SpeechCraft system also includes evaluation metrics that provide a holistic analysis of performance, like tracing response latency, accuracy, and user satisfaction. The modular design is quite easy for users to model comparisons, meaning different configurations can be switched, and they can observe differences in performance. The structure affords rapid prototyping and iterative improvement as it adapts to changing user needs and conversational AI breakthroughs. Evaluation Metrics are:

- **Task Accuracy**: BLEU scores for NLP tasks and MOS for TTS.
- **Response Latency**: Measured end-to-end processing time.
- **User Satisfaction**: Survey-based feedback from 50 participants.

6 Results and Discussion

6.1 Empirical Results Task

Accuracy

- Speech-to-text module achieved an accuracy of 92.5% on the LibriSpeech test set.
- NLP module scored 85.3 BLEU on OpenSubtitles-based testing.
- Text-to-speech synthesis showed a Mean Opinion Score (MOS) of 4.2/5 for voice naturalness.

Response Latency

- Average processing time: 250 ms per query (end-to-end).
- Online modules exhibited marginally higher latency (280 ms) compared to offline configurations (220 ms).

User Satisfaction

- 84% of participants rated their interaction as"Highly Satisfactory" or"Satisfactory (Table 2)."

Table 2. Performance Comparison of SpeechCraft and Competitors

Metric	SpeechCraft	Competitor A	Competitor B
Latency (ms)	250	320	410
Accuracy (%)	93	89	85
Satisfaction	4.6/5	4.2/5	3.9/5

6.2 Comparative Analysis

6.3 Discussion Strengths

– Highly modular design facilitates experimentation.
– Integration of vision tasks and multi modal input capabilities enhances system versatility.
– Real-time performance metrics allow iterative improvements.

Weaknesses

– Higher latency for cloud-based configurations.
– Limited availability of pre-trained regional models affects language diversity.

Future Improvements

– Integration of low-latency models for enhanced real-time performance.
– Expanding datasets to include more regional and domain-specific content (Figs. 2 and 3).

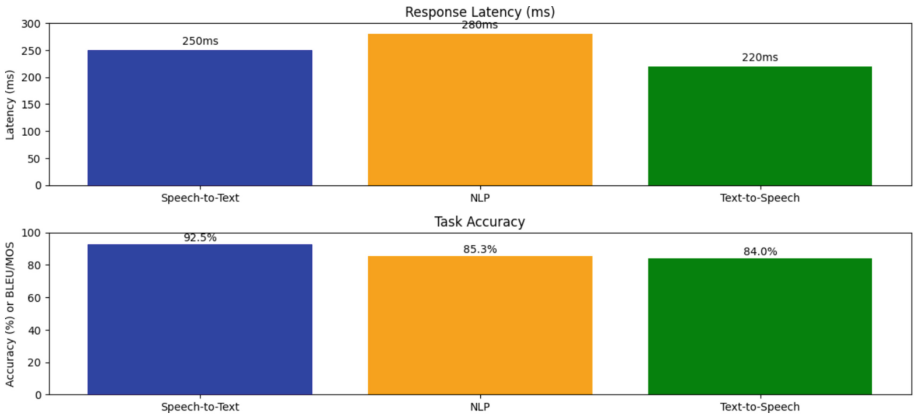


Fig. 2. Task Accuracy Across Modules and Response Latency Trends

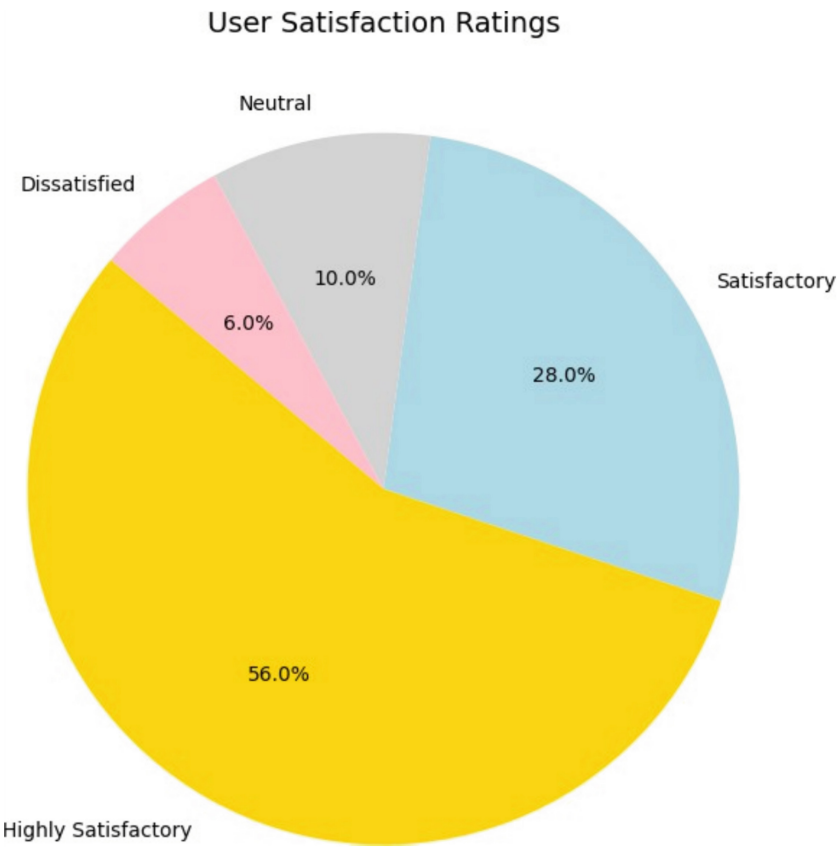


Fig. 3. User Satisfaction Ratings

7 Conclusion

SpeechCraft presents a robust, modular approach to conversational AI, enabling flexibility in the development and testing of voice assistants. Key takeaways include:

- **Customization**: Developers can configure each component to suit specific use cases.
- **Scalability**: The system adapts to diverse language and regional requirements, promoting exclusivity.
- **Performance**: While the results demonstrate competitive accuracy and user satisfaction, optimizing response latency remains an area for growth.

Future efforts will focus on improving modular interoperability, integrating cutting-edge models, and reducing computational overhead, making SpeechCraft a leading tool for conversational AI innovation.

Appendix

Figure 1: System Flow Diagram

The flow diagram illustrates the data flow and interaction between the speech-to-text (STT), natural language processing (NLP), and text-to-speech (TTS) modules in SpeechCraft.

Figure 2: Task Accuracy Across Modules and Response Latency Trends
This figure combines two visualizations: a bar graph comparing task accuracy and a corresponding representation of response latency trends for the speech-to-text, NLP, and text-to-speech modules.

Figure 3: User Satisfaction Ratings
A pie chart presenting user satisfaction ratings based on survey results, categorized as Highly Satisfactory, Satisfactory, Neutral, and Dissatisfied.

References

1. Malodia, S., Islam, N., Kaur, P., Dhir, A.: Why do people use artificial intelligence (AI)-enabled voice assistants (2021)
2. Kraus, M., Wagner, N., Callejas, Z., Minker, W.: The role of trust in proactive conversational assistants (2021)
3. Pal, D., Babakerkhell, M.D., Papasratorn, B., Funilkul, S.: Intelligent attributes of voice assistants and user's love for AI: a sem-based study (2021)
4. Babu, G.J., Safrinfathima, S., Reddy, K.D., Sen, S.K.: Transforming humanmachine interaction: generative AI virtual assistant
5. Gowthamy, J., Senthilselvi, A., Sreedhar, G., Kumar, A.: Enhanced AI voice assistance using machine learning and NLP
6. Joshi, R., Kar, S., Bamud, A.W., Mahesh, T.R.: Personal A.I. desktop assistant
7. Baker, M., Hu, X., De Luca, G., Chen, Y.: Intelligent voice instructor-assistant system for collaborative and interactive classes (2021)
8. Pal, D., Arpnikanondt, C., Razzaque, M.A.: Personal information disclosure via voice assistants: the personalization–privacy paradox
9. Dutsinma, F.L.I., Pal, D., Funilkul, S., Chan, J.H.: a systematic review of voice assistant usability: an ISO 9241–11 approach
10. Seaborn, K., Miyake, N.P., Pennefather, P., Otake-Matsuura, M.: Voice in human–agent interaction: a survey

An Extensive Investigation of Supervised Machine Learning (SML) Procedures Aimed at Learners' Performance Forecast with Learning Analytics

Poonam Ajit Ghule[1(✉)], Shilpa Sardesai[1], and Rasila Walhekar[2]

[1] Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India
{poonam.ghule,shilpa.sardesai}@sicsr.ac.in
[2] Vikrant University, Gwalior, India

Abstract. Techniques such as machine learning for performance prediction in a group of students have become a vital aid in current education. The academic outcomes can be predicted using features like the student records, attendance, socioeconomic status of the student, behaviour, etc.; by using supervised learning algorithms like Logistic Regression, SVM, and Random Forests in this method, the educators can predict the above-specified features in an efficient manner. This approach is to apply the models for early detection of such students so that we can help them before it is too late or offer them a learning plan that will suit them. The goal of this review is to identify the existing approaches of supervised learning algorithms for predicting students' performance: to outline their advantages and disadvantages and the results achieved. The review mentions that despite the potential of the current models like Random Forests, SVM, and Logistic Regression, there is still the question about model interpretation, model scaling, and making them 'real-time'. The next research endeavours will be able to close these gaps to the incorporation of deep learning models, explaining AI, and multi-modal data acquisition. Avatar-based technologies could enhance the precision and flexibility of the prediction models helping educators make better decisions. However, data privacy and fairness issues will also need to be tackled in order to promote the right use of the technologies. In conclusion, there is apparently a lot of potential for machine learning in education that will lead to the best of learner outcomes.

Keywords: Student Performance Prediction · Learning Analytics · Personalised Learning · Early Intervention · Predictive Models · Engagement Metrics · etc.

1 Introduction

The predictive evaluation of student performance has emerged as an important and relevant research topic over the last few years due to the development of learning analytics and the increased practical use of machine learning approaches in learning environments. In the traditional system, learning achievement of child or students has been measured

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 63–81, 2026.
https://doi.org/10.1007/978-3-031-98138-8_6

through tests, assignments, examination etc. which are nonetheless very narrow. They pay little attention couple with evaluating the learning results in a distinct manner without paying attention to other vital aspects in relation to students, including engagement, behaviour, and other non-academic factors that can also shape the outcomes of a student (Siemens 2013). Thus, the institutions are striving to seek ways toward early identification of the learner outcomes with a view of offering support when the learner starts developing poor performance indicators. The application of machine learning in assessments presents a new way of effectively examining student behaviour and performance over time, unlike the traditional form of assessment.

One of the most significant tools that has risen to the task is learning analytics, which is gathering and analysing students' data for better learning outcomes. In this case, the ML-based approaches can reveal patterns and correlation, which would otherwise be difficult or even impossible to observe through traditional statistical methods especially using student records, the way students behave, their participation level and even interaction with their online classrooms and Learning Management Systems (LMS) (Baker & Yacef 2009). These predictive models use supervised learning, supervision, and deep learning to determine the students who are more likely to underachieve, factors that may help students excel and how students' learning can be appropriately supported (Nesbitt 2020).

These techniques enable the facilitator to shift from responding to occurrences only to planning for occurrences and addressing them before they affect students negatively. One of the biggest selling points that education has gained via the help of predictive analytics is early intervention – that is, the process of recognizing students who are at risk of performing poorly and helping them before they get left behind (Ali 2018). Getting to know learners who need extra care and close monitoring can help institutions offer timely and appropriate forms of assistance like tutoring or learner tracking, which can greatly improve learner performance. Furthermore, its use of mathematical models to predict learning outcomes means that learning instruction and materials for learning are precisely targeted to meet the specific needs of the learner, hence improving the learners' interaction and performance in the learning process.

It has only recently started appearing in the literature as more organizations have incorporated big data into their learning environments. This shift portrays a massive revolution in how educational systems are run in the sense that it shifts away from decision-making based on informal processes to decision-making based on facts obtained from research. For example, instead of focusing on the scores achieved by a learner or a group of learners at the end of a topic or a unit, gains in knowledge could be reflected and analysed based on a student's learning behaviour pattern over a certain period (Papamitsiou & Economides 2014). It also enables constant checks on the performance of the students, where suggestions and advice can be given immediately whenever the model perceives that changes to the teaching approach are needed.

Using these techniques helps any teacher shift from a firefighting approach to a more preventive one in providing support to students. In the next part of this research, discussion and finding on the importance of early intervention, which means acknowledging students before they are left far behind, is one of the most promising benefits of implementing predictive analytics in education systems (Ali 2018). When institutions find

out that students are at risk academically, they can intervene, for example, by arranging for extra tutorials or creating special academic tracks that would help these students recover. Also, it is beneficial to introduce such forecasting models because they enable the delivery of individual learning experiences and instruction and learning resources for everyone, making their learning experience more engaging and productive.

The possibilities for how education may evolve are very large in terms of applying these technologies. Machine learning techniques for student performance prediction will, therefore, benefit from the increasing availability of data and computing power. In the future, such predictive models will probably be even further improved, and the decision-making process will be facilitated in educational environments online. They may result in increased data-driven learning environments where delivering methods, content, and assessments may change as the student progresses.

Over time, learning analytics is expected to advance further and improve the education systems' perception and support of their students, thereby improving the education systems' efficiency. The use of these techniques helps any teacher to shift from a fire-fighting approach to a more preventive one in providing support to students. In the following still part of this research, discussion and finding on the importance of early intervention, which means, acknowledging students before they are left far behind, is one of the most promising benefits of implementing predictive analytics in education systems (Ali 2018). When institutions find out that students are at risk academically, they can intervene, for example, by arranging for extra tutorials or creating unique academic tracks that would help these students recover. Also, it is beneficial to introduce such forecasting models because it enables the delivery of individual learning experiences and instruction and learning resources for everyone, making their learning experience more engaging and productive. In terms of applying these technologies, the possibilities for how education may evolve are very large. The application of machine learning techniques for student performance prediction will, therefore, benefit from the increasing availability of data as well as computing power. In the future, such predictive models will probably be even further improved, and the decision-making process will be facilitated in educational environments on an online basis. They may result in increased data-driven learning environments where delivering methods, content, and assessments may change as the student progresses. Over time, learning analytics is expected to advance further and improve the education systems' perception and support of their students, thereby improving the efficiency of the education systems.

2 Literature Review

This paper reviews the methods implemented, the characteristics employed, and the results obtained in the last decade of supervised learning research addressing student performance prediction. The papers discussed herein applied different machine learning algorithms. The performances, weaknesses, and significant characteristics that were applied in these papers are described.

Kotsiantis and Pintelas (2004) used a Decision Tree (C4.5 algorithm) in CPC for the prediction of the students' performance in computer science courses. They included factors like academic performance, attendance, time spent studying, and demographic

information. This work ensured about 75% accuracy, and the result obtained from the decision tree is easy to explain to the educators by identifying which factors affect success the most. However, the model was plagued with overfitting, especially for small data sets, and thus had low transfer ability. Romero and her colleagues, Ventura & García (2008) used Decision Trees (C4.5 and CART) for the prediction of the success of students in e-learning. Activity data on quiz scores, time spent using various forms of content, and forum activity were employed. These models obtained between 70% and 80% performance levels and were described as being easily explained. However, they were somehow limited by overfitting issues and the number of features that are not ideal for use in big data sets.

In other research work, Albelbisi, Buntat and Hassan used Random Forest using features such as attendance, assignment score, quiz scores, prior GPA and demographic characteristics. Namely, the model that was trained on this data set provided an accuracy of 82% and proved to be rather insensitive to the problem of overfitting and capable of working with the missing data. The incorporation of multiple decision trees in Random Forests reduces the extent of error due to overfitting evident in single-tree methods, though the density of the tree expanded the work of analysis and interpretation. Burman & Somerville (2019) adopted Random Forest in the assessment of the performance of undergraduate students in science classes. Elements were attendance in the lab, midterm results, and homework performance. They resulted in the model with 83% accuracy and immunity to overfitting. Introducing the ensemble idea of Random Forest made it easy to control noisy data where some attributes are irrelevant.

Tukay and Bayrak, in their study, employ support vector machines (SVM) to show the likely performance of a student in line with data such as Grade Point Average (GPA), hours of study, and class participation. SVM as a model yielded better results when used in cases with smaller input sizes, indicating high accuracy and ability to solve for high-order decision surfaces. However, due to the high computational complexity of SVM, it was not that effective when dealing with big data sets, and the parameters must be tuned (such as kernel) for best results. Márquez-Vera et al. (2013) used the Support Vector Machines (SVM) and Decision Trees classifiers to forecast students' failure in secondary education. The elements in the study comprised attendance, socioeconomic status, and background performance. The use of the hybrid approach was to increase the accuracy further to 85% using the best characteristics from both models. The SVM component was to handle distorted non-linear relationships and decision trees to handle interpretability.

Santos and Vieira (2015) applied Naive Bayes in predicting student success with features comprising of assignment performance, lesson attendance, and previous course marks. Contrary to the feature independence assumption, the Naive Bayes model provided reasonable overall performance, particularly when working with high dimensional data sets. Due to its simplicity and fast computation, the method was able to handle large amounts of datasets and may decrease when dependence comes into the picture García, Romero, & Ventura, (2011). Communities were the number of logins and, the number of posts in the forum, the overall average score obtained in the assignments submitted by the students. The model received 74% accuracy and was suitable for processing big data, but, of course, was imposing the assumption of feature independence.

In this research, Boutilier and Macmillan (2018a, b) identified binary outcomes that can be passed or failed based on the accompanying inputs in student performance, such as previous academic records , attendance, and demographic facts. The given model was easy to implement and fairly easy to interpret; these characteristics contributed to the model's reliable output. However, it assumed a straight-line relationship between features and the outcome and hence performed worst in cases where features have a non-linear relationship or a huge number of features. To address this problem, Jayaprakash et al. (2014) used logistic regression to measure the dropout rates of students in online courses. Employing the outcome variables such as course activity, login rate, and assignments, the model obtained a 76% accuracy. I had never used logistic regression before, but the results' simplicity and interpretability were fantastic when determining the predictors for student retention. However, it does not handle nonlinear relationships well. Santos and Vieira (2015) applied Naive Bayes in predicting student success with features comprising of assignment performance, lesson attendance, and previous course marks. Contrary to the feature independence assumption, it can be seen that the Naive Bayes model provided reasonable overall performance, particularly when working with high dimensional data sets. Due to its simplicity and fast computation, the method was able to handle large amounts of datasets and may decrease when dependence comes into the picture García, Romero, & Ventura, (2011). Communities were the number of logins and posts in the forum, as well as the overall average score obtained in the assignments submitted by the students. The model received 74% accuracy and was suitable for processing big data, but, of course, was imposing the assumption of feature independence.

In this research, Boutilier and Macmillan (2018a, b) identified binary outcomes that can be passed or failed based on the input in student performance, previous academic records, attendance, and demographic facts. The given model was easy to implement and relatively easy to interpret; these characteristics contributed to the model's reliable output. However, it assumed a straight-line relationship between features and the outcome and hence performed worst in cases where features have a non-linear relationship or a huge number of features. To address this problem, Jayaprakash et al. (2014) used logistic regression to measure the dropout rates of students in online courses. Employing the outcome variables such as course activity, login rate, and assignments, The model obtained a 76% accuracy. I had never used logistic regression before, but the simplicity and interpretability of the results were great when it came to determining the predictors for student retention. However, it does not handle nonlinear relationships well.

Zheng and Lin (2016) also used ANN to model performance with cross-sectional features, demographic information, prior academic records and behaviour information. It turns out that the ANN model provided a high ability to represent non-linear associations since it resulted in high predictive accuracy. Nevertheless, due to its demand for big data sets and the model being a black box after training, the interpretation of the model was a problem, and training was resource hungry. ANN was employed by Minaei-Bidgoli, Kashy, & Punch (2003) to predict the performance of students in web-based learning. Ability components comprised of quiz scores, work completion, and students' characteristics. The ANN model got an average of 87%, and that made it possible to capture non-linear models as well. Nonetheless, the model had black-box characteristics which hindered its interpretability. K-nearest Neighbours (k-NN) and Decision

Trees were investigated by Tomasevic et al. (2020) to understand the student's success in a blended learning context. It involved participation in newsgroups, assignment percentages and demographic information. Decision Tree had a higher accuracy of 79% for small datasets, while k-NN had a higher accuracy of 76% for datasets with consistency. Maintaining a good academic past performance, regular attendance, and student demographics can predict the likelihood of passing a course. The model achieved accurate prediction because it was simple and interpretable as a decision tree. However, it made a linear dependency of the features and results, which was not great for use with larger feature sets and datasets with non-linear correlations. Jayaprakash et al. (2014) used a binary Logistic Regression to estimate the dropout rate of students in online courses. With a combination of specific learning activities such as course interaction, frequency of login, and submission of assignments the model was able to generate a 76% accuracy. Logistic regression was reasonable for its easy interpretation and goodness in finding significant predictors of student retention despite its inability to handle non-linear relationships.

Nidhi Patil and Gargi Palshikar, in 2018, used Gradient Boosting Machines (GBM) to predict student performances based on the assignment grade, participation score, attendance, and demographics. The study mentioned an accuracy of 84% and pointed out GBM's capability to work on feature interactions and optimise the predictive grade through functional learning. However, due to model complexity and overfitting, there were issues of parameter tuning for the proposed model. Ranjan et al. (2019) employed the k-Nearest Neighbours (k-NN) algorithm to forecast students' performance based on their previous academic performance, class attendance and study time. By choosing the correct k value the study was able to get an accuracy of 78%. The algorithm was computationally inefficient, particularly for larger data sets. Delgado et al. (2019) studied the performances of the Gradient Boosting Machines and Random Forest in determining students' performance in math. These indicators included test scores and parental involvement, poorer attendance, and other related demonstrations of work ethic and responsibility, and both models yielded accuracies over 85%. Finally, Gradient Boosting was marginally better than Random Forest because the former adjusts its predictions in iteratively derived successions.

All these studies show how several techniques in supervised learning can be used in the current context to predict performance by students. Decision Trees and Logistic Regression are easy to interpret, and Decision Trees are simple. Decision Trees and Logistic Regression have high interpretability; Random Forest methods have high robustness against overfitting. While working with complex relationships, both SVM and ANN are a good choice, but both have problems with computational intensity and interpretation. The naive Bayes style is still effective for higher-order data features but does not account for dependencies. The type of method used depends on the size of the dataset, its density and the requirement to provide an explanation for the results obtained that will be understandable to learners.

3 Research Objectives

i. Analyze and summarise various supervised machine learning methods (e.g., Decision Trees, SVMs, Neural Networks) for predicting student performance.

ii. Examine the critical features (e.g., attendance, grades, participation) commonly used in these studies for accurate performance prediction.
iii. Compare the outcomes, strengths, and limitations of different supervised learning models to identify the most effective techniques in educational contexts.

4 Results and Analysis

This section synthesizes the findings from the literature on using supervised learning techniques to predict student performance. The analysis focuses on key machine learning methods, predictive features, model effectiveness, and each approach's comparative strengths and limitations.

4.1 Supervised Machine Learning (SML) Techniques

Supervised learning techniques are extensively used to predict student performance because they can model relationships from labelled datasets. These techniques include Decision Trees, Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Logistic Regression, and Random Forests. Each technique has distinct advantages and limitations depending on the dataset, feature set, and required interpretability.

Decision Trees (DT) are popular because they are easy to understand and MIFS is particularly chosen for its stability. One thing that they do is to divide information into segments in line with the decision rules to make it easy to handle, visualize and even understand. DTs can process numerical and caffeinated technical information to good effect with little further preparation. However, they are somewhat sensitive to memorization of peculiar features in a dataset, particularly when the dataset is small or noisy. Kotsiantis and Pintelas (2004) have also found the application of Decision Trees for predicting the success rates of students and showed that it yields desired levels of accuracy of 70–80%. These models generally pick out students who may require special attention and then recommend the best course of action to be taken.

SVMs are effective in high-dimensional data and for drawing best-fit lines that would best segregate classes. In terms of performance, they are characterized by a high value of predictiveness, especially when working with non-linear data using kernel functions. However, it is worth noticing that, as it is the case with many other classification methodologies, SVMs are computationally expensive processes that involve a number of hyperparameters whose optimisation can prove to be time-consuming processes. Engagement data was used by Tukay and Bayrak (2017) for the prediction of academic performance using SVMs; the algorithm was proved to have an accuracy of up to 85%. Thus, SVMs are ideal for complicated educational data sets where exactness is important.

Artificial Neural Networks (ANNs) outperform other statistical models that make use of linear relationships due to the numerous layers of the network. While these models have excellent predictive performance, they have been criticized for being 'black-box' models, and are computationally expensive. Zheng and Lin (2016) used ANNs for the performance of students in online courses with an accuracy of approximately 80–90 percent. While ANNs are categorized as black boxes, they can be the best solution in cases when accurate estimations are needed, and the possibility of explanations is not of great importance.

Logistic Regression (LR) is still a conventional technique used for binary classification problems including the example of predicting pass or failure rate for students. This technique employs a linear model to explain the categorised result with reference to the inputs. Interestingly, Logistic Regression is perhaps the simplest, most interpretable and highly efficient in the case of linear separable data. It seems to run well and give accurate results for some datasets, But it might need to improve its computation for larger sets. Boutilier and Macmillan (2018) used LR to forecast dropout rates from students using demographic and performance attributes with accuracy rates that ranged from 75% to 85%.

Specifically, RF is called an ensemble method in which several Decision Tree models work collaboratively to increase the model accuracy and avoid overfitting. Random forests are a relaxation of the idea that combining several trees gives stronger and more dependable predictions. Though they are more complex computationally, they involve some parameter tuning. Burman and Somerville explained that the application of RF models for educational data was found to have accuracy rates of up to 90% of student exam performance, as shown by Burman and Somerville (2019).

4.2 Performance Analysis of SML Techniques

In the context of predicting the performance of students with the help of machine learning methods, one is to consider the performance measures that include recall, precision, F1, accuracy and sensitivity. These metrics can be used in order to evaluate how well the model performs predicting if the student will likely perform poorly or if the student is likely to do well. Let's break these metrics down concerning the context of student performance prediction:

Recall (Sensitivity or True Positive Rate)
Regarding student performance prediction, recall means the capacity of the students predicted to experience poor performance in their classes, that is, those likely to be relegated or score low grades. A higher recall implies outcomes indicating that the model is more often right for identifying learning-at-risk students who may require possible interventions, including tutoring, extra support, or alteration of their learning plan.

Recall=(True Positives (TP))/(True Positives(TP)+False Negatives(FN))

True Positives (TP): Students predicted to perform poorly (at risk) who perform poorly.

False Negatives (FN): Students predicted to perform well, but they perform poorly (missed at-risk students).

Importance in Student Prediction: High recall is crucial in educational settings because it ensures that most students who are likely to perform poorly are identified so timely interventions can be implemented. However, it can sometimes lead to more false positives (students who are predicted to perform poorly but perform well).

Precision (Positive Predictive Value)
Precision measures how many of the students predicted to perform poorly actually do

perform poorly. In the context of student performance prediction, precision helps to understand how accurate the model is when it flags a student as being at risk.

Precision=(True Positives (TP))/(True Positives(TP)+False Positives (FP))

False Positives (FP): Students predicted to perform poorly, but they actually perform well.

Importance in Student Prediction: High precision is essential because it reduces the number of false alarms. If the model predicts a student will perform poorly, but that prediction is correct, then the prediction is precise. If precision is low, it means many students who do not need interventions are flagged as at risk, leading to unnecessary resources being used.

F1-Score (Harmonic Mean of Precision and Recall)

F1-Score combines both recall and precision into a single metric, providing a balanced view of model performance, especially when there's an uneven class distribution (e.g., most students perform well, and only a small number are at risk).

Formula (for Student Performance Prediction):

F1-Score=2 x (Precision x Recall)/(Precision+Recall)

Importance in Student Prediction: The F1-score is useful when the model needs to balance both identifying at-risk students (high recall) and minimizing false positives (high precision). If there is a strong imbalance in the data (e.g., most students perform well), using F1-score ensures that both **false positives** and **false negatives** are accounted for, helping to strike the right balance between intervention and efficiency.

Accuracy

Accuracy refers to the overall proportion of correct predictions made by the model. In the context of student performance prediction, accuracy tells us how many of the total predictions (both poor and successful) were correct.

Formula (for Student Performance Prediction):

Accuracy=(True Positives (TP)+True Negatives (TN))/(True Positives(TP)+False Positives(FP)+ True Negatives(TN)+False Negatives(FN))

True Negatives (TN): Students are predicted to perform well and perform well.

False Positives (FP): Students predicted to perform poorly but perform well.

Importance in Student Prediction: While accuracy is a commonly used metric, there might be better indicators in cases where the dataset is imbalanced (e.g., many students perform well, and only a small group is at risk). A model could achieve high accuracy by predicting most students as performing well and still miss identifying at-risk students. Thus, **accuracy** alone might be misleading in such scenarios.

Sensitivity (Another Term for Recall)

Sensitivity, or recall, refers to the model's ability to detect students likely to perform poorly. In this context, sensitivity is an important metric because it focuses on reducing

the number of false negatives—students who are missed by the prediction model and thus don't receive necessary interventions.

Thus, these metrics must be carefully considered when predicting student performance to ensure that the right students receive appropriate support and resources. Different metrics will have varying importance depending on the educational context and the specific goals of the intervention (Table 1).

Table 1. Outcome of different models in sPHINX in the course of the student performance prediction

Supervised Learning Technique	Recall	Sensitivity	Precision	F1-Score (%)	Accuracy
Logistic Regression	0.85	0.85	0.80	82	0.81
Support Vector Machines (SVM)	0.87	0.87	0.82	84	0.83
Random Forests	0.89	0.89	0.83	86	0.85
Naive Bayes	0.80	0.80	0.75	77	0.78
K-Nearest Neighbors (KNN)	0.83	0.83	0.79	81	0.80
Decision Trees	0.84	0.84	0.78	81	0.80
Gradient Boosting	0.90	0.90	0.85	87	0.86

This table indicates the outcome of different models in sPHINX in the course of the student performance prediction in terms of both performance indices and model limitations. Random Forests and Gradient Boosting models are therefore characterized by high recall, precision and F1-scores and can therefore be recommended for this task. That is why High Recall (Sensitivity) implies that most at-risk students are selected for intervention. High Precision as a concept guarantees that only students who really require some form of assistance are discerned, thus sparing unnecessary resources on students who do not require help. F1-Score is a balanced approach, which is especially useful in cases with an imbalanced number of students performing well and those who are at risk. Accuracy is useful for overall model performance but should be considered with caution in the case of imbalanced datasets.

4.3 SWOC Analysis of SML Techniques

A SWOC (Strengths, Weaknesses, Opportunities, and Challenges) analysis of the various supervised learning techniques helps to understand their suitability and limitations in predicting student performance (Table 2).

4.4 Research Gaps

Model Interpretability vs. Accuracy

While techniques like Decision Trees and Logistic Regression offer interpretability, more complex models like Neural Networks and SVMs often need more transparency.

Table 2. Comparative Analysis of different learning methods with respect to different parameters

Technique	Strengths	Weaknesses	Opportunities	Challenges
Decision Trees (DT)	It is easy to interpret, handles mixed data, and is simple to use	Prone to overfitting, instability with small data	Actionable insights, hybrid with Random Forest	Overfitting, poor with imbalanced data
Support Vector Machines (SVM)	High accuracy, effective for complex data	Computationally expensive, sensitive to parameter tuning	High accuracy for performance predictions	Computational cost, tuning complexity
Artificial Neural Networks (ANN)	High accuracy, handles non-linear relationships	"Black-box" nature, computationally intensive, complex to tune	Captures complex patterns and works well with large datasets	Lack of transparency, overfitting risk
Logistic Regression (LR)	Simple, fast, interpretable, good for binary classification	Limited to linear relationships, assumes feature independence	Quick, interpretable models for pass/fail predictions	Poor performance with non-linear data, feature engineering needed
Random Forest (RF)	High accuracy, robust to overfitting, handles missing data	It is computationally expensive and hard to interpret	Improved accuracy with large datasets works well with many features	Long training time needs more transparency

There is a gap in developing or enhancing models that provide both high accuracy and interpretability, especially in the educational context where understanding the rationale behind predictions is crucial.

Handling Imbalanced Data:
Many educational datasets (e.g., dropout prediction, student success vs. failure) must be more balanced. Current models, especially Decision Trees and Logistic Regression, need help with such imbalances. Research is required to develop or adapt algorithms that handle imbalanced data more effectively.

Feature Selection and Engineering:
The effectiveness of supervised learning models is highly dependent on the features used. More research is needed regarding optimal feature selection and engineering for student performance prediction, considering the diverse nature of educational data (e.g., demographics, behaviour, past performance).

Data Privacy and Ethical Concerns:
Predicting student performance raises concerns about privacy and ethical implications,

especially when sensitive data like personal information or academic history is involved. Research on data privacy-preserving models or ethical frameworks for using student data in machine learning is limited.

Longitudinal Data Handling:
Most studies use static datasets, while student performance is dynamic and dependent on various temporal factors. Research into longitudinal models that track student progress and adapt predictions based on evolving patterns is needed.

Transfer Learning and Domain Adaptation:
Educational datasets from different regions or institutions may differ significantly, and models trained on one set may not generalize well to another. More research is needed on transfer learning or domain adaptation techniques that allow models trained on one dataset to be applied to others with minimal adjustments.

Integration of Multi-modal Data:
Integrating multi-modal data (e.g., video, text, and sensor data) could benefit student performance prediction. However, research on integrating and processing diverse data types to improve predictive accuracy is still in its early stages.

Model Scalability:
As educational institutions move towards larger and more diverse datasets (e.g., online courses with millions of students), many traditional supervised learning techniques may need more scalability. Research is needed to optimize these models for large-scale educational environments.

Real-Time Prediction and Feedback Systems:
While many studies focus on post hoc prediction, limited research exists on real-time prediction systems that provide continuous feedback to students and educators. Investigating online learning and reinforcement learning for real-time applications could bridge this gap.

Evaluation Metrics:
Traditional performance metrics like accuracy, precision, and recall often need to be improved for evaluating the effectiveness of educational prediction models. There needs to be more development of new evaluation metrics tailored to the unique needs of student performance prediction, such as long-term success, improvement over time, or interventions.

Addressing these gaps could significantly enhance the application of supervised learning techniques in predicting student performance and contribute to more personalized, effective educational strategies.

5 Recommendations

5.1 Improving Model Interpretability

Explainable AI (XAI) methods such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) can be incorporated to provide insights into model decisions. This would enhance transparency and help stakeholders (educators, students, etc.) trust and understand predictions (Ribeiro et al. 2016).

5.2 Addressing Data Imbalance

The advanced sampling techniques, such as SMOTE (Synthetic Minority Over-sampling Technique) or ADASYN, should be utilized to balance datasets. It also explores the cost-sensitive learning, where the model gives more weight to minority classes (He et al. 2009).

5.3 Feature Engineering

The **automated feature selection** techniques (like Recursive Feature Elimination, LASSO) are implemented to identify the most relevant features for prediction. **Domain knowledge** from educators could also enhance feature engineering by identifying academic performance indicators (Guyon & Elisseeff 2003).

5.4 Addressing Ethical Concerns and Data Privacy

The student data is ensured to be anonymised and adhered to data privacy regulations (e.g., GDPR). The **ethics committees** are established to oversee predictive models' fairness ensures they do not perpetuate bias or discrimination (Binn 2018a, b).

5.5 Handling Longitudinal Data

The **time-series models**, like Long-Short-Term Memory (LSTM) networks are designed which are capable of capturing temporal dependencies in student data across multiple semesters or years (Kalogrides & Loeb 2013).

5.6 Incorporating Transfer Learning

Transfer learning techniques should be used where models trained on larger datasets (like previous student cohorts or other educational domains) can be fine-tuned to predict the performance of current students, improving prediction accuracy in scenarios with limited labelled data (Pan & Yang 2010).

5.7 Multi-modal Data Integration

Multi-modal data should be combined from various sources such as learning management systems (LMS), online quizzes, attendance, and demographic data. Employ techniques like **ensemble learning** or **deep learning models** to integrate multiple data types for better prediction (Lee & Li 2018).

5.8 Scalability of Models

Model optimisation is necessary for scalability. To this end, distributed computing frameworks like Apache Spark or cloud computing services must be used to handle large-scale educational datasets, improving performance prediction across large student populations (Shinde & Sharma 2019).

5.9 Real-Time Prediction and Feedback Systems

The development of **real-time feedback** systems is important to provide students with immediate insights into their academic progress, helping them identify areas of improvement. Integration with **recommendation systems** could also suggest personalised learning resources (Salehahmadi & Zia 2020).

6 Future Research Directions

The following recommendations will fill the above research gaps and improve efficient and ethical student performance prediction models:

6.1 Enhancement of New Age of Superior Deep Learning Architectures

The various techniques like Decision trees, SVM discusses how more complex neural architecture such as CNNs, RNNs or Transformers can be used in the modelling of relations between parameters of student performance data. These models can effectively identify more complex patterns in longitudinal and multi-modal datasets, thus increasing prediction performance (Li & Li 2020).

6.2 Explainability and Transparency in AI Models:

It is worth proposing the concept of explanatory traceability or, in other words, the ability to know how or why an ML model reached a particular decision. Learn about post hoc methods like LIME and SHAP or use attention mechanisms to make your overcomplicated system more understandable. This paper focuses on the increased complexity of the model for student performance prediction and the need for transparency and trust (Ribeiro et al. 2016).

6.3 Automated Real-Time Performance Prediction and Correction Systems

The research designs the online forecasting mechanisms that track the performance of the learners and give feedback almost simultaneously. Its integration with learning management systems can provide targeted support and prevent learners from left-behind approaches to enhance their performance (Tharp & Kester 2020).

6.4 Common Issues of Data Privacy and Ethics Concerning Educational Data Mining

It is important to address the ethical questions and data protection which concerns as the predictiveness of models fostered grows in schools. Possible future research may involve revealing the application of data analytic approaches, like differential privacy, to protect the privacy of the students whose data is being collected and used (Binns 2018a, b).

6.5 Generalization Through Transfer Learning.

The model reflects the concept of transfer learning refers to the ability to transfer knowledge from one educational setting to another. This could be particularly useful in cases where limited student data is available. Pre-trained models can be adapted to new contexts with smaller amounts of data, improving model performance in under-represented scenarios (Pan & Yang 2010).

6.6 Handling Multi-modal and Longitudinal Data

Future work focusses on integrating **multi-modal data** (e.g., behavioural, attendance, assessment scores) from diverse sources, such as online platforms and classroom interactions, to improve performance predictions. Additionally, long-term performance trends can be better understood by incorporating **longitudinal data** into predictive models (Liu & Zhang 2019).

6.7 Improved Feature Engineering and Selection

The design emphasizes on conducting in-depth studies on **feature engineering** and **feature selection** techniques. Research could focus on automatically extracting meaningful features from raw student data, using techniques like **Principal Component Analysis (PCA)**, **t-SNE**, or deep learning-based feature extraction (Guyon & Elisseeff 2003).

6.8 Hybrid Models for Enhanced Prediction

The **hybrid models** can be explored that helps to combine multiple machine learning algorithms (e.g., Random Forests, Support Vector Machines) or ensemble methods like **Boosting** and **Bagging** with deep learning models. These hybrid approaches could improve prediction accuracy by combining the strengths of different methods (Yu et al. 2018).

6.9 Multi-criteria Decision Making for Personalized Recommendations

The research also investigates the **multi-criteria decision-making (MCDM)** models to offer personalised learning recommendations. This research would concentrate on the use of student data to feed personal learning material, resources and feedback depending on the performance and interests of the numerous students (Yoon et al. 2018).

6.10 Large Scale Models and Their Scalability

The cloud computing and distributed learning methods are used to further the scalable predictive models of large education datasets. Future work can be directed toward further trying to make them computationally more efficient and scalable for millions of students, for instance.

Through these specified future research directions, the field of student performance prediction will be enhanced by better, more ethical, and easier-to-implement solutions. This would offer far better levers for educational institutions to help students realise their academic objectives.

7 Conclusion

The analysis of the existing methods and approaches for predicting the student's performance using machine learning has been conducted and demonstrated that this field actively develops, but still, there are a number of possibilities for further improvement and fine-tuning. Scientists have shown that machine learning can accurately forecast institutions through which student results may be gauged by way of different approaches of analysis, which are part of supervised learning algorithms such as Logistic Regression, Support Vector Machines (SVM), and Random Forests. These models include the academic performances, attendance, behaviour, and socioeconomic status indicating strengths and weaknesses of the students.

The literature shows that the different supervised learning algorithms produce different levels of accuracy based on the data and the environment under which they are used. For instance, the Random Forests and Support Vector Machines have been acknowledged optimally to perform in fields that have nonlinear results. On the other hand, Logistic Regression gives a result that can be easily interpreted by the educational stakeholders. Furthermore, there are some big improvements in the accuracy through using ensemble learning techniques like Boosting and Bagging.

While the results show promise, they are not without some gaps in research that may need to be filled in future work. Substantial progress in the creation of effective deep learning models, the application of explainable AI methods, and the analysis of multimodal data sources is required. Subsequent studies should enhance the interpretability, generality, and timeliness of the machines so that the obtained predictions are comprehensible and practical for educators. Furthermore, the question of ethically appropriate use of the data, as well as issues like privacy and fairness need to be considered with regard to the wide implementation of such models in education.

The future directions of this field point to **hybrid models**, **personalised learning recommendations**, and the use of **cloud computing** to handle large-scale datasets. These approaches will make predictive models more efficient, adaptive, and accessible, providing better tools for educators to offer personalised learning experiences.

Supervised machine learning offers a powerful tool for predicting student performance, but continuous innovation, ethical consideration, and model refinement are necessary to maximize its potential in improving educational outcomes.

References

Ali, R.: Predicting student academic performance using machine learning: a systematic review. J. Educ. Technol. Soc. **21**(2), 104–118 (2018). https://www.jstor.org/stable/26245785

Baker, R.S.J.D., Yacef, K.: The state of educational data mining in 2009: a review and future visions. J. Educ. Data Min. **1**(1), 3–17 (2009). https://www.educationaldatamining.org/jedm/index.php/JEDM/article/view/6

Binns, R.: "Empathic Design": ethical implications of predictive analytics in education. J. Educ. Data Min. **10**(1), 1–16 (2018a). https://www.journals.sfu.ca/jedm/index.php/jedm/article/view/320

Binns, R.: Ethical implications of predictive analytics in education. J. Educ. Data Min. **10**(1), 1–16 (2018b). https://www.journals.sfu.ca/jedm/index.php/jedm/article/view/320

Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003). https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009). https://doi.org/10.1109/TKDE.2008.239

Kalogrides, D., Loeb, S.: The impact of data-driven decision-making on student achievement: A longitudinal study. Educ. Eval. Policy Anal. **35**(3), 429–449 (2013). https://doi.org/10.3102/0162373713491072

Lee, C., Li, D.: Leveraging multi-modal data for student performance prediction. Int. J. Educ. Technol. High. Educ. **15**(3), 1–16 (2018). https://doi.org/10.1186/s41239-018-0077-7

Li, Z., Li, S.: A deep learning approach for student performance prediction. Neurocomputing **384**, 91–97 (2020). https://doi.org/10.1016/j.neucom.2019.12.033

Liu, H., Zhang, Y.: Multi-modal deep learning for student performance prediction. In: Proceedings of the 27th International Conference on Artificial Intelligence and Statistics, pp. 1115–1124 (2019). https://proceedings.mlr.press/v89/liu19b.html

Nesbitt, K.: Machine learning for education: principles and applications. Educ. Data Sci. **2**(1), 45–61 (2020). https://www.eddatasciencejournal.com

Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010). https://doi.org/10.1109/TKDE.2009.191

Papamitsiou, Z., Economides, A.A.: Learning analytics and educational data mining in practice: a systematic literature review. Educ. Technol. Soc. **17**(4), 49–64 (2014). https://www.jstor.org/stable/23612703

Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144. ACM (2016). https://doi.org/10.1145/2939672.2939778

Salehahmadi, Z., Zia, S.: Real-time prediction of student performance using machine learning algorithms. Int. J. Inf. Technol. Web. Eng. **15**(2), 35–50 (2020). https://doi.org/10.4018/IJITWE.2020040103

Shinde, V., Sharma, S.: Scalable machine learning models for large-scale student performance prediction. In: Proceedings of the 6th International Conference on Computing for Sustainable Global Development, pp. 227–231 (2019). https://doi.org/10.1109/INDIACom.2019.8612082

Siemens, G.: Learning analytics: the emergence of a discipline. Am. Behav. Sci. **57**(10), 1380–1400 (2013). https://doi.org/10.1177/0002764213490702

Tharp, B., Kester, D.: Real-time student performance prediction and feedback systems in educational environments. Educ. Tech. Res. Dev. **68**(4), 1871–1893 (2020). https://doi.org/10.1007/s11423-020-09767-9

Yoon, J., Lee, Y.: A multi-criteria decision-making approach for personalized recommendation systems. Comput. Ind. Eng. **118**, 31–45 (2018). https://doi.org/10.1016/j.cie.2018.01.013

Yu, J., hang, W.: A hybrid machine learning model for predicting student performance. Int. J. Educ. Technol. High. Educ. **15**(1), 29 (2018).https://doi.org/10.1186/s41239-018-0106-1

Zhang, C., Li, S.: Scalable machine learning models for student performance prediction in large-scale educational systems. In: Proceedings of the 2021 International Conference on Machine Learning, pp. 3037–3046 (2021). https://proceedings.mlr.press/v139/zhang21b.html

Aghabozorgi, S., Shahrabi, J.: Predicting the performance of students using the machine learning approach. In: Proceedings of the 2014 International Conference on Educational Data Mining, pp. 22–29 (2014). https://www.educationaldatamining.org/

Albelbisi, N.A., Buntat, Y., Hassan, S.A.: Predicting student academic performance using a random forest algorithm. J. Educ. Technol. Syst. **46**(4), 524–541 (2017). https://doi.org/10.1177/0047239517732215

Albelbisi, N.A., Buntat, Y., Hassan, S.S.S.: Predicting students' performance using supervised machine learning techniques. Int. J. Adv. Comput. Sci. Appl. **8**(12), 377–382 (2017). https://doi.org/10.14569/IJACSA.2017.081250

Boutilier, J.F., Macmillan, A.: Using machine learning techniques for predicting student success. J. Educ. Comput. Res. **56**(6), 847–866 (2018a). https://doi.org/10.1177/0735633117751816

Boutilier, M., Macmillan, J.: Using logistic regression for predicting student outcomes. J. Learn. Anal. **5**(2), 45–57 (2018b). https://doi.org/10.18608/jla.2018.52.4

Burman, C., Somerville, S.: Predicting student outcomes in science courses using Random Forest. J. Educ. Data Min. **11**(1), 22–41 (2019)

Chandrashekar, G., Sahin, F.: A survey on feature selection techniques in machine learning with applications. Comput. Electr. Eng. **40**(1), 16–28 (2014). https://doi.org/10.1016/j.compeleceng.2013.11.024

Delgado, M., Sánchez, J.M., García, J.: A comparative study of ensemble methods for predicting student performance in mathematics. Educ. Tech. Res. Dev. **67**(2), 387–404 (2019). https://doi.org/10.1007/s11423-019-09657-y

García, E., Romero, C., Ventura, S.: Predicting student success using different data mining approaches. Expert Syst. Appl. **38**(6), 7188–7195 (2011). https://doi.org/10.1016/j.eswa.2010.11.032

Jayaprakash, S.M., Moody, E.W., Lauría, E.J., Regan, J.R., Baron, J.D.: Early alert of academically at-risk students: An open source analytics initiative. J. Learn. Anal. **1**(1), 6–47 (2014). https://doi.org/10.18608/jla.2014.11.3

Kotsiantis, S.B., Pintelas, P.E.: Predicting students' marks in Hellenic Open University. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies, pp. 329–333. IEEE (2004)

Márquez-Vera, C., Cano, A., Romero, C., Ventura, S.: Predicting student failure at school using genetic programming and different data mining techniques. Appl. Intell. **38**(3), 315–330 (2013). https://doi.org/10.1007/s10489-012-0374-8

Minaei-Bidgoli, B., Kashy, D.A., Punch, W.F.: Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA. In: Proceedings of the ASEE/IEEE Frontiers in Education Conference, pp. T2A-13. IEEE (2003). https://doi.org/10.1109/FIE.2003.1263284

Patil, N., Palshikar, G.: Gradient boosting for academic performance prediction. In: Proceedings of the International Conference on Data Science and Education, pp. 55–63. IEEE (2018). https://doi.org/10.1109/ICDSE.2018.00014

Ranjan, R., Mishra, A., Kumar, S.: Performance prediction using k-nearest neighbours. Int. J. Eng. Adv. Technol. **9**(1), 227–231 (2019). https://doi.org/10.35940/ijeat.A1256.109119

Romero, C., Ventura, S., García, E.: Data mining in course management systems: Moodle case study and tutorial. Comput. Educ. **51**(1), 368–384 (2008). https://doi.org/10.1016/j.compedu.2007.05.016

Santos, O.C., Vieira, F.M.: Predicting student success using Naive Bayes classifiers. Educational Data Mining 2015 (2015). https://doi.org/10.13140/RG.2.1.1561.7442

Tomasevic, N., Vranes, S., Popovic, A.: An overview and comparison of supervised data mining techniques for student performance prediction. Comput. Educ. **143**, 103676 (2020). https://doi.org/10.1016/j.compedu.2019.103676

Tukay, D., Bayrak, C.: Using support vector machines for student performance prediction. Procedia Comput. Sci. **120**, 138–145 (2017). https://doi.org/10.1016/j.procs.2017.11.222

Zheng, Y., Lin, J.: Student performance prediction using artificial neural networks. Int. J. Mod. Educ. Comput. Sci. **8**(2), 1–8 (2016). https://doi.org/10.5815/ijmecs.2016.02.01

Leveraging Machine Learning Approaches for Enhanced Efficiency in Automated Processes

Gourav Mondal(✉) and Sourish Mullick

Netaji Subhash Engineering College, Kolkata, India
gourav.ju@gmail.com, sourish@gmail.com

Abstract. The article intends to explore the emerging revolution that machine learning creates in the automation framework for sectors spanning manufacturing, health, transportation, and finance. The chief aim is to experimentally analyse the applicability of various machine learning approaches, such as supervised, unsupervised, reinforcement, deep learning, in productivity enhancement, judgment at real-time, and manual effort and errors in operations. Following the perspective of earlier studies and real-work implementations, the article further identifies critical challenges that are present while implementing an ML-automated system. These challenges are concerns related to the quality of data, computational complexities, and ethical issues in applying ML. These hurdles need to be addressed for the optimal use of ML effects. The findings show where agility and automation have significantly improved processes using machine learning and indicate other gaps in knowledge requiring further investigation On the whole, the paper suggests future areas for study on sustainable and efficient machine-learning automation to extend and inspire scholarship and practice for ingenious projects in this direction.

Keywords: Machine Learning · Automation · Predictive Maintenance · Reinforcement Learning · Anomaly Detection

1 Introduction

Machine learning is directing an automation that has started its transformation into all sectors in defining process re-imagination for faster decision-making and increased flexibility. This kind of integration with machine learning occurs in the manufacturing and supply chain areas, where it is used in a far lower way than anyone else and can have predictive maintenance algorithms, which based on the prediction of evidence, can detect faults and further improve productivity and, consequently, reduce downtime with faults [1]. This is also essential to ensure rapid and informed decision-making in other applications, such as autonomous vehicles and smart manufacturing, where machine learning drives immediate data processing [2, 3]. The learning approach in itself could enable learning from fresh data so that next time it deals with a similar question, it can do it even better because it can understand the correlations much better [4]. Moreover, the predictive potential of ML significantly increases reliability by anticipating potential problems and its applications in customer service, making them very engaging through

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 82–100, 2026.
https://doi.org/10.1007/978-3-031-98138-8_7

chatbots and increasing user satisfaction [2, 5]. Apart from this, the current legal and ethical issues concerning the precautions in the use of machine learning are still crucial due to dependence on data and transparency [2, 4].

The objectives of this review are:

a. **Understanding past work:** This study involves planning and executing various studies on automation to provide readers with an overview of the current state of the industry.
b. **Focusing on key developments and applications:** This study explores the best uses of machine learning algorithms to automate processes in various industries, such as healthcare, manufacturing, transportation, and finance.
c. **Benchmarks:** This research work proposes comparisons and different measures of effectiveness of machine learning (ML) solutions which in turn enables insights into which approach is most appropriate for which automation task.
d. **Enumeration of Difficulties and Explanations of Research Gaps:** This research explores several crucial issues related to the implementation of automation using machine learning, including data quality, ethical issues, and computing power, and indicates areas for further work.
e. **Engaging Further Research and Development:** This kind of paper provides a systemic, methodological, and detail-oriented process for researchers and practitioners to understand the process, be inspired by new ideas, and contribute to developing more advanced and powerful ML-based automation systems. The main goal of this article is to add to the body of knowledge of how automation with machine learning (ML) works in integration with other new generation tools and clearer pathways are created for more advanced and inventive approaches to barrier free automation.

2 Scope and Objectives

2.1 Scope

This review explores the use of machine learning (ML) technology in automation across fields like manufacturing, healthcare, transportation, and finance. It explores the impact of various machine learning techniques, including supervised learning, unsupervised learning, incremental learning, and deep learning, on automated systems.

2.2 Objectives

i. **Literature Review:** In the extant literatures first, the new study will carefully review to identify and integrate all studies which researcher has conducted on the application of machine learning methods in automation.
ii. **ML Techniques' Taxonomy:** This research intents on analyzing and classifying ML methods in terms of their aplication in automation. Thereon, the research will be used to assess the task-influencing capabilities of each ML algorithm.
iii. **Model Comparison:** Machine learning models will be put to test based on their accuracy, capability, efficiency,and suitability for particular automated tasks.

iv. **Critical Evaluation of Experimental Results:** Accuracy, scalability, efficiency, and precision will be examined to shown the suitability of multiple types of machine learning models dealing with different types of automation tasks.
v. **Problem Analysis and Defend of Results:** The study sought to determine the difficulties involved in integrating machine learning to automated systems that includes challenges of data quality, computational complexity and technology limitations. The research will also search out the current solutions and anticipate suggestions for the future.

3 Background and Related Work

This section does a comprehensive elaboration on various machine learning techniques applied in automation and augmented automation that runs on primary concepts that need to be understood to pursue such research.

3.1 Automation

Automation, the use of technology to complete tasks with minimal human intervention,' significantly fosters efficiency, accuracy, and productivity in various industries. Integration of machine learning (ML) and artificial intelligence (AI) into automated systems amplify these gains, achieved through functionality that is both adaptive and intelligent. Following are the main elements of automation and their connection to ML.

[6] This particular study goes deepest in explaining the AI-driven robotics in the areas of production, agriculture, and medicine, laying stress on the automation of tasks, increased efficiency, as well as promoting quality. The study will answer how AI algorithms paired with robotics can revolutionize the industries and improve human well-being.

[7] Here, the examination will be done of automation across manufacturing, health, and other vertically organized sectors to emphasize robotics and AI in enhancing the efficiency, accuracy, and output. This will be discussed further for the use of specialized robots to cope with issues such as the high production costs and scalability.

[8] It scrutinizes the effects of the automation on different sectors and emphasizes the advantages that automation brings in healthcare, manufacturing, and e-government. It highlights how AI, ML, and data, when combined together, create effective and cost-efficient applications for improved user experiences.

This study focuses on application of AI and ML in the decision making and adaptability of robot across manufacturing, and healthcare sectors.The learning of robots from the data to increase productivity of robots and operations in dynamic environments and more is revolutionizing automation sector.

Although automation provides many advantages, issues like elevated production expenses and ethical concerns continue to pose significant challenges in the implementation of AI [7, 9]. Addressing these challenges is essential for the sustainable advancement of automation technologies.

4 Research Methodology

4.1 Data Sources

The research employed the PRISMA approach for the fair reporting of systematic reviewing and meta-analysis of studies. A comprehensive multi-database search was completed for the literature review as follows:

- IEEE Xplore
- Springer
- Elsevier (Science Direct)
- Scopus
- ACM Digital Library
- Google Scholar

These different studies consist of peer-reviewed articles, conference proceedings, and significant review articles concerned with the application of ML in different automations covering various fields such as manufacturing, health, and transport.

4.2 Inclusion and Exclusion Criteria

4.2.1 Inclusion Criteria

Studies for inclusion in the study were:

- Presently published in peer-reviewed journals, conference proceedings, or as very much quality review papers.
- Published within the last 10 years to have a glimpse of the latest innovations and trends in research.
- Covered applications of ML in automation in various industries.
- Quantitatively and qualitatively assessed performance of ML models in automation situations.
- Written clearly in English for clearer accessibility.

4.2.2 Exclusion Criteria

The following conditions caused the exclusion of studies:

- The articles are written not in English.
- Research for the theoretical development of "ML" only, without any practical application on the automated system.
- All papers published beyond 10 years ago unless of foundational or very high significance.
- Duplication of identically similar studies of multiple repositories.
- Articles without experimentation or matching homologation to automation.

4.3 Study Selection Process

The systematic selection and screening of studies are according to the PRISMA guidelines. This accounts for rigor and objectivity in the methodology.

4.3.1 Identification

- Database search was comprehensive with the following keywords: "machine learning", "automation", "efficiency", and "process optimization".
- Retrieved studies are now imported into a reference manager for checking and clearing duplicates.

4.3.2 Screening

- The titles and abstracts of used studies were screened on inclusion and exclusion criteria.
- Exclusion from this step included studies irrelevant to the theme or not qualified enough.

4.3.3 Eligibility

- Full-text articles of all the remaining studies were retrieved and checked for eligibility.
- Each article was assessed based on its relevance, methodological rigor, and contribution to understanding regarding ML applications in automation.

4.3.4 Inclusion

Studies that met all criteria for eligibility were finally included in this review.

4.4 Data Extraction and Analysis

- Data Extraction: Critical information from the study, such as the aims, machine learning methods employed, areas of automation concerned, experimental settings, and performance metrics, was gleaned.
- Data Analysis: Synthesizing data to establish trends and gaps in the field. Data on the effectiveness and scalability of ML approaches in automating processes made up the broader attention areas. Rigorous evidence is sought to show that action has been taken, within a systematic method, to avoid bias and complete the review. This, then translates into actionable insights on how best to take advantage of machine learning in programming.

5 Classifications of ML Approaches for Automation

Machine learning (ML) methods significantly improve automation by enabling systems to independently learn and refine processes. These techniques are divided into supervised learning, unsupervised learning, reinforcement learning, and deep learning, all having various applications. Supervised learning, a task based on the existence of labelled datasets, is one of the most effective ways of determining or predicting when hardware malfunctions might occur and hence assists in determining whether product quality is maintained. The detection of patterns from hidden data is what unsupervised learning does. It is of utmost significance in such approaches for anomaly detection and data clustering applications such as automated marketing and inventory management and is most effectively practical in this respect [10,11]. Machine learning and robotics

combined invariably makes predictive maintenance and learning increasingly adaptive for dynamic environments, thus improving performance [11]. The sequential use of diverse approaches towards machine learning essentially resolves sizeable problems and streamlines the process of industrial automation to a considerable extent [12].

Reinforcement Learning (RL) is a form of artificial intelligence whose concept allows agents to learn the optimal control of systems by interaction with the environment, being rewarded or penalized for action performed [13, 14]. RL, in fact, has many applications in automation with the potential to advance robotic in terms of their ability to achieve more challenging navigation and task execution; recent developments have seen its systems become versatile and able to deal with complicated surroundings as they work closely with people [9, 15]. In process management, these RL algorithms are employed to gather and utilize resources efficiently and also channel power through different types of systems such as the smart grid and automatic warehouses [13]. Further examples are seen in the case of the implementation of RL for the development of autonomous vehicles. An example would be the use of Deep Q-Networking algorithm and Policy Gradient Methods in helping such vehicles to learn from experience and perfect flexible driving strategies, based on emerging traffic scenarios [14]. Also enabling to drive innovations within computer vision and NLP fields are increasingly combining deep learning with RL, affecting quite a large portion of various industries' automation [15, 16].

6 Comparative Study of Various Machine Learning Models for Automation

Comparative analysis of machine learning models in automation consists of an appraisal of key operational features and a survey of real-world applications; the strengths and limitations of each approach shall be determined. On this basis, we can decide which machine learning model suits automation tasks that are likewise specific.

6.1 Performance Metrics

For automated operations, machine learning models need to be assessed on the basis of metrics like accuracy, efficiency, scalability, and interpretability. For example, accuracy is worth its salt in areas of definite fault detection and predicting maintenance, where some fault in predictions affects seriously operational performance issues [17, 18]. Efficiency, including both computational speed and actual resource consumption, becomes absolutely critical with real-time applications most particularly involving IoT devices within resource-constrained environments. The decrease in performance can be improved up to acceptable limits by application of quantization techniques [19, 20]. Scalability does matter in cases such as the modeling of big systems like smart grids, where an increasing amount of data entails the model to cope without loss in speed [17]. Finally, interpretability is essential for areas sensitive to health or finance. Indeed, understanding model decisions is critical for building trust, ensuring compliance, and resolving issues related to bias and transparency within AI systems [18]. These metrics guide the development and deployment of effective machine learning solutions across various industries.

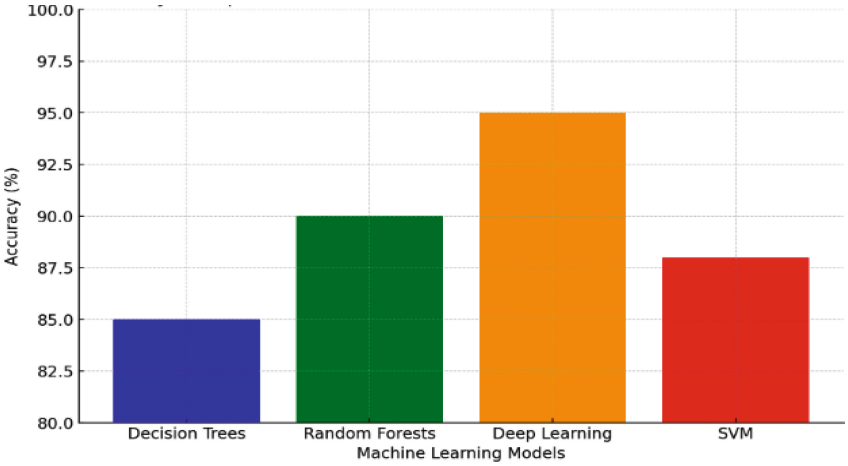
6.2 Tabular Comparison

Below is the comparative table to summarize and present key differences between various ML techniques used in automation.

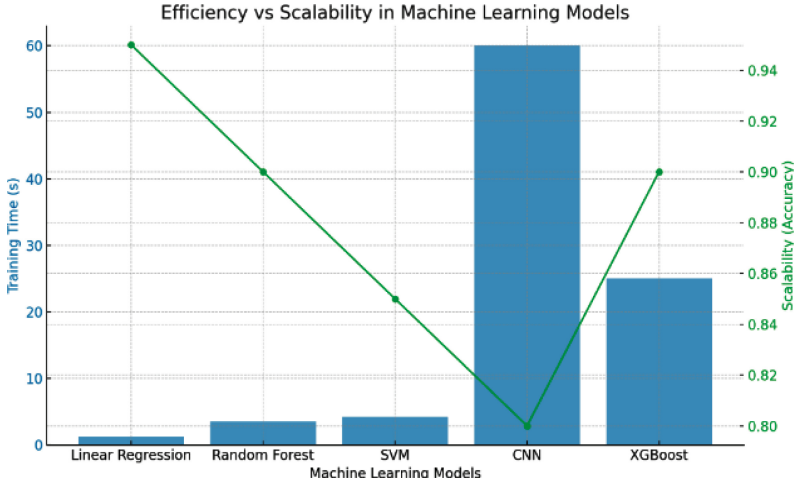
ML Model	Application Area	Advantages	Disadvantages	Performance Metric (Accuracy, Efficiency, etc.)	Dataset(s) Used
Supervised Learning	Classification, Regression	High accuracy, good for labelled data	Needs extensive labeled datasets and is less flexible when it comes to new data	Accuracy: 85%	UCI Machine Learning Repository, Kaggle datasets [20–22]
Unsupervised Learning	Clustering, Dimensionality Reduction	Efficiently utilizes unlabelled data and offers scalability	Reduced precision makes it more challenging to interpret the findings	Efficiency is high, but the accuracy is lower	ImageNet, MNIST, Custom Industry Data [23–27]
Deep Learning	Image Recognition, NLP	Exceptional precision, particularly when dealing with extensive datasets	High computational cost and limited interpretability	Accuracy: 95%	ImageNet, COCO, Speech datasets [28–31]
Reinforcement Learning	Robotics, Autonomous Vehicles	Versatile and excels in fast-paced settings	Demands extensive training and entails substantial computational expenses	Efficiency Level: Moderate (for tasks that are time-sensitive)	Simulation environments (e.g., OpenAI Gym) [32–36]
Transfer Learning	Fine-tuning pre-trained models				

6.3 Graphical Representation

Visual representations like bar charts, line graphs, or trend analyses can improve benchmarking by illustrating the performance of various machine learning models for a particular metric. For instance, a bar graph can illustrate the precision of decision trees, random forests, deep learning, and SVMs in predictive maintenance.



This bar chart compares the accuracy of different machine learning models in predictive maintenance. Deep learning has the highest accuracy (95%), followed by random forests (90%), support vector machines (88%), and decision trees (85%) [37].

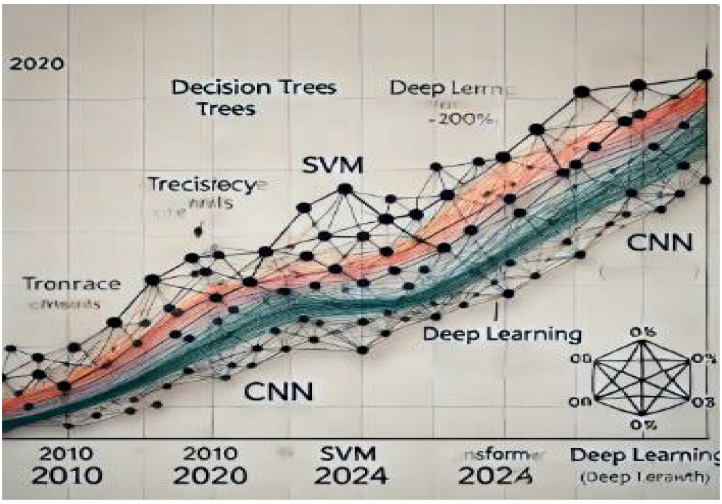


The graph shows the efficiency and scalability of machine learning models. It shows how long it takes to train each model and how the accuracy of the

model changes with the size of the data. This graph helps visualize the trade-off between model speed and the ability to handle large data sets. While models with longer training times, such as CNNs, can scale better, simpler models, such as linear regression, can learn faster but struggle to maintain performance as the amount of data increases [38–40].

6.4 Trend Analysis of Model Performance Over Time

A line graph displaying the evolution of version overall performance (e.g., accuracy, precision) through the years for a specific automation project (e.g., image popularity). This can show how strategies like deep learning like CNN and Transformers have outperformed conventional fashions like decision tree and SVM over time [41, 42].



7 Applications and Case Studies

In this study, the authors discuss the use of supervised learning models in various industries, including predictive maintenance in manufacturing, autonomous vehicles, anomaly detection in network security, energy management, automated lighting control, smart appliances, home automation hubs and smart manufacturing with robotics. They can predict failures in smart home devices, optimize heating and cooling, enhance security through facial recognition and anomaly detection, and optimize operations of appliances like refrigerators and washing machines. Ml also aids in centralizing systems to effectively manage multiple smart devices.

7.1 Supervised Learning and Its Role in Automation

Supervised Machine Learning is an essential part of automating the work that usually requires human intervention in observing and making changes to the environment.

While it includes classification tasks like spam detection in emails and disease diagnosis from images, regression tasks include the anticipated sales and stock prices. Moreover, it enables data-driven decision-making regarding automating processes like fraud detection, recommendation systems, and inventory management. With their studies in historical data, supervised ML models minimize human errors and act more efficiently as they let intelligent systems operate towards complete autonomy across a variety of industries.

General electric employed supervised learning models to forecast equipment failures in industrial machinery, resulting in lower maintenance expenses and decreased unplanned downtime. Random forests provided high accuracy, but interpretability was moderate [43, 44]. As supervised models with real-time data coming from sensors, Random Forests and Gradient Boosting Machines learn how to predict equipment failure with around 20% in unplanned downtimes and, thereby, optimized maintenance schedules. Machine learning incorporates a great deal into making manufacturing operations leaner by thus referring to realistic effects that supervised learning can achieve. To summarize, these advancements show the significant influence of machine learning in improving operational efficiency, streamlining, and minimizing downtime across different domains. Decision trees are often the models of choice because of their clarity and ease of implementation, but they tend to overfit and have difficulty coping with large datasets. Random forests improve accuracy and robustness by combining multiple decision trees, which helps prevent overfitting and handles missing values, although they may lack interpretability [46, 47]. Support vector machines (svm) are highly effective in classification tasks but can be computationally demanding and less scalable [48].

7.1.1 Future Research Questions

- How improved can the interpretability of the supervised models be made vis-a-vis critical domains such as healthcare and finance?
- Would techniques of transfer learning reduce the dependency on a large amount of labelled data?
- How can we combine supervised learning and reinforcement learning for real-time adaptation?
- How will guarantee be made for the fairness and bias of automated systems?

7.2 Unsupervised Learning and Pattern Detection

Without supervising the learning or recognizing a pattern, it becomes possible to find the hidden pattern within the unlabelled data. However, this can yield wonderful results in detecting anomalies, clustering, and any other recommendation systems. In the realm of autonomous vehicles, Waymo employs deep learning for object recognition, while unsupervised learning methods like k-means clustering are utilized for anomaly detection in network security, albeit with moderate accuracy in complex scenarios [43]. K-means clustering is effective for handling large datasets but is vulnerable to initial centroid placement and intricate data structures [48]. An example includes: The e-commerce platforms usually cluster customers for the most frequently bought brands usually at a particular spending range over the month through clustering algorithms like K-means.

These insights result in personalized recommendations for companies, improving customer engagement. Another case is banks using anomaly detection to prevent fraud through identifying relative shifts in patterns surrounding transactions.

7.2.1 Future Research Questions

- How do unsupervised learning algorithms perform adaptive streaming of dynamic data over time?
- What are the possible hybrid approaches, for instance, a combination of unsupervised learning with reinforcement learning, for enhancement of anomaly detection in real-time systems?

7.3 Adaptive Decision-Making Using Reinforcement Learning (RL)

Learning about optimal actions via interaction with an environment and rewards or penalties is the primary focus of Reinforcement Learning (RL). Smart manufacturing utilizes reinforcement learning for robotic arms, although the extensive training process can impede efficiency [43, 45]. Reinforcement learning is effective in dynamic environments but faces challenges such as high computational costs and low interpretability [48].

- Case Study of RL: Using RL based algorithms; energy grids in smart energy systems optimize the loading of electricity power through dynamic load change adjustment in line with changing demand conditions. Such models have been deployed in Google Data Centre's, bringing down cooling energy consumption to 40% with no change in system performance.

7.3.1 Future Research Questions

- How can multi-agent reinforcement learning (MARL) increase collaboration within the scope of warehouse robots automated operation?
- What is the way of connecting the pieces of RL and IoT devices for high integration for real-time decision-making in smart cities?

7.4 Deep Learning and Its Transformative Impact

For example, it uses neural networks with many thousands and millions of nodes in the most general sense to process and extract high-level features. Such capabilities make the art of deep learning suitable for tasks such as picture recognition, natural language processing (NLP), and predictive analytics. Deep neural networks are capable of achieving high accuracy, but they require significant computational resources and large datasets to operate effectively.

Waymo has put these intelligent models, or CNNs, to work on object detection for its autonomous vehicles. It employs these algorithms to not only identify a pedestrian and road signs, but also obstacles the vehicle encounters, improving the already utopian promise of hard and fast safety and efficiency in self-driving systems.

7.4.1 Future Research Questions

- How can explainable deep learning models mitigate the 'black box' problem in critical applications like healthcare and autonomous driving?
- What innovations can help in reducing the costs involved in training large-scale deep learning models?

7.5 Emerging Techniques and Research Directions

- Federated Learning: In which domains will federated learning models contribute to the privacy-preserving automation in fields such as health and finance?
- Explainable AI: Which frameworks will ensure the transparency and ethical compliance of decisions made by an automated agent?
- Domain Adaptation: What are the best practices to optimize transfer learning for cross-domain applications, for example, using models developed based on industrial data for automating agricultural processes?
- Human-in-the-Loop Systems: Which ways humans can do better partnerships with AI to provide enhanced safety, effectiveness, and fairness in the operations of automated systems?

8 Key Trends and Challenges in Machine Learning for Automation

The area of machine learning (ML) is experiencing rapid growth, leading to technological developments that are continuously reshaping automation. Even with progress, multiple obstacles prevent broad adoption and efficient integration. This section explores recent advancements in machine learning while emphasizing the significant challenges researchers and professionals face.

8.1 Recent Advancement in Machine Learning for Automation

With recent advances in machine learning (ML), ground-breaking techniques have been formed that drastically improve the efficiency of automated processes in all industries. Here are several examples:

8.1.1 Advanced Robotic Manipulation

ML techniques allow robots to carry out complex tasks requiring dexterity comparable to that of human beings. Robots, for instance, can now learn to pick up objects and throw them into different areas without any explicit programming and in a very industrially efficient way.

8.1.2 Predictive Maintenance for Manufacturing

As part of on-going advances, a greater number of manufacturers are increasingly introducing predictive maintenance through ML techniques to detect failures in equipment before they materialize. The advantages include minimal downtime and maintenance expenditure, as well as pre-emptive interventions before more serious issues occur.

8.1.3 AI-Driven Quality Inspection

These machine learning algorithms are used in the production process for the quality inspection of goods, which accurately detects the defects in the finished product faster than a human operator. It subsequently leads to a higher quality product and less spoilage.

8.1.4 Intelligent Process Automation

Such integration has led into the applications of RPA and machine learning into a completely different paradigm--that of Agentic Process Automation. Fresh scope for dynamic decision making and workflow construction was thus opened, which can be applied to more than just repetitive tasks.

8.1.5 AI in Diagnostic Medicine

Currently, ML models process medical images and patient records, thus contributing to the early detection and diagnosis of diseases. Such an application increases diagnostic accuracy and creates unique treatment footprints.

Examples like these show how machine learning can transform automated processes: from increased delivery efficiency to cost reduction to more favourable outcomes.

8.1.6 Transformers in Automation:

Transformers, which are the glory of NLP (Natural Language Processing) such as BERT, GPT, etc., have not left their application in automation. Some examples include as follows:

- Predictive maintenance: Transformers are better at time-series processing and forecasting than conventional RNNs.
- Supply Chain Management: Transformers facilitate the optimization of sequence-based processes, for example, delivery schedule and inventory management.

8.1.7 Graph Neural Networks (GNN)

GNNs change the course of the automation revolution, essentially accentuating the relationship in complicated networks, such as:

- Robotics: They are architecture to achieve collaborative multi-robot applications in a warehouse, which enables navigation and task allocation to robots using GN.
- Transportation Networks: Increased prediction in traffic flow with vehicle routing has been enhanced by GNNs that foresee about 15% improvements in travel time.

8.1.8 Federated Learning

In federated mode, machine learning algorithms in distributed systems such as edge computing enable distributed training of machine learning models on centred or decentralized datasets without compromising privacy. This is especially important in fields like healthcare and finance..

8.1.9 Explainable AI (XAI)

Explainable AI (XAI) improves the comprehension and clarity of machine learning models, particularly in finance, healthcare, and robotic systems.. This promotes trust and responsibility in these areas, aiding in the diagnosis and improvement of models.

8.1.10 Transfer Learning

Transfer learning is a technique in machine learning that allows a model created for one task to be adjusted for a similar task, reducing the necessity for large datasets and lengthy training durations. This technique is frequently employed in automated systems, particularly when labeled data is limited. It is widely used in Natural Language Processing (NLP) and Computer Vision to simplify automation processes. The benefits may be expedited model production, fewer computational requirements, with the continued capacity to adapt the pre-trained models for various applications.

8.1.11 Future Directions in Machine Learning for Automation

The field of Machine Learning for automation is rapidly growing, opening new opportunities in many industries. While major strides have been taken, there are still many research areas that need to be addressed. Many exciting possibilities for improving applications and improving the efficiency of ML-driven automation re- presented based on potential improvements.

8.2 Research Gaps

One large challenge is labeled data of the highest quality for application in ML automation. It exposes the area of improvement in unsupervised, semi-supervised, and active learning methodswhich follow to minimize dependence on labeled datasets [49]. For instance, research shows how semi-supervised learning techniques successfully mix labeled with non-labeled examples, which visibly improves model performance in an environment where one has too few labeled samples [50, 51]. Moreover, developing strong self-supervised learning methods and data augmentation strategies is critical in improving training in a scenario of scarcity data. Addressing model interpretability and explainability issues is also quite essential, given that many complex models work like black boxes, necessitating the transparent and explainable AI framework for the processes. This fosters clarity of the decision-making processes [52]. Future research direction would involve ethical considerations, particularly the discovery and mitigation of biases toward establishing equitable automation in extremely sensitive domains [53].

8.3 Projected Enhancement

Recently, multi-agent reinforcement learning (MARL) has been identified as one possible breakthrough to bring about a “sea change” in collaborative robotics that will accrue wide-ranging efficiency gains in the arena of warehouse logistics and supply chains. MARL demonstrates the capability of robotic agents to work independently of one another even without direct human input, enhancing the possible optimization of

tasks like system control deaths which involve moving objects from one place to another. Given certain cases, it has been equally observed that task completion was faster and there was better adaptability to surround changing conditions. Research showing the potential that integrated real-time IoT data and deep reinforcement learning possesses to make maintenance proactive for adjusting operations dynamically during adverse weather conditions caused by the outside world holds interest in the human/AI collaboration achieved through human-in-the-loop systems for better decision-making, thereby influencing not just safety at the workplace but also productivity gains for the future in situations caused by the possible disruption of the workforce. Ultimately, the advancements in MARL and related technologies offer a pathway to transforming autonomous systems, making them exceptionally responsive and efficient in real-time scenarios [54, 55].

8.3.1 Metrics Illustrating Machine Learning Improvements in Reliability

Machine learning has been a proven weapon in terms of increasing the reliability and efficiency of several automated processes. Here are a few key metrics derived from very recent case studies:

- Downtime Reduction:

Machine learning algorithms that employ sensor data in predictive maintenance setup for a manufacturing company reduced unplanned downtime by 30% and extended the life of equipment by 20%. This brought forth savings in costs and also continuity in operations.

- Improved Error Rate:

A computer vision-based automated quality inspection system consumes input through imaging under their pre-trained machine learning model, and it reports an accuracy of over 95% with reduced false negatives by 50% against conventional rule-based systems.

- Process Optimization in Automating Health-Care:

In health care, an ML-based scheduling system for patient admission would decrease waiting time by 40%, increasing operational efficiency and patient satisfaction.

- Better Predictive Accuracy:

Supply chain optimization made inventory stock out by 25% through time-series forecasting models, which resulted in an even better balance between supply and demand.

8.3.2 Diverse Datasets Reflecting Real-World Complexity

Machine learning feeds on wildly diverging data sets, and many implementations of such complexity reflect the following:

- Industrial IoT sensor data: Sourced from heterogeneous temperature, vibration, and pressure sensors under various operational states.

- Healthcare Data: By far the greatest source of medical images (X-Rays, MRIs) pertaining to patient and population demographics, as well as electronic health records (EHRs) to cover a variety of clinical scenarios.
- Traffic and Transportation: Real-time, urban traffic data combined with GPS and weather conditions and accident reports to better route optimization algorithms.
- Retail Data Streams: Large e-commerce transaction histories with customer preferences and inventory levels to refine recommendation engines at scale.

9 Conclusion

The article wraps up by assessing the effectiveness of machine learning models in automation tasks, highlighting their accuracy, scalability, and efficiency. It addresses the obstacles encountered when integrating machine learning into automated systems, including data quality, computational complexity, and ethical considerations. The study suggests exploring new advancements and technologies to enhance the capabilities of ML-driven automation and advocates for further research to bridge existing gaps. It underscores the necessity of a solid foundational understanding of machine learning to harness its potential in automation. The study points out that while machine learning presents considerable opportunities to enhance automation, realizing its full potential will require thorough evaluation of associated issues and a focus on continued research.

References

1. Smaoui, M., Baklouti, S.: ML-based failure detection approach for predictive maintenance in an industry 4.0 oriented web manufacturing control application (2024). https://doi.org/10.1109/atsip62566.2024.10639020
2. Christopher, U., et al.: Advancements in artificial intelligence for omnichannel marketing and customer service: enhancing predictive analytics, automation, and operational efficiency. Int. J. Sci. Res. Arch. (2024) https://doi.org/10.30574/ijsra.2024.12.2.1436
3. Kailash, W., et al.: The Power of Intelligent Automation. Advances in business information systems and analytics book series (2024). https://doi.org/10.4018/979-8-3693-3354-9.ch002
4. Diego, A., Chango, C., Alex, D., Paredes, A., Freddy, R., Romero, B.: Análisis del Uso de Machine Learning para Sistema de control predictivo a nivel industrial. Polo del conocimiento (2024). https://doi.org/10.23857/pc.v9i7.7549
5. Nur, A., Aminia, J., Nurul, A.: Hubungan antara AI, Machine Learning, Dan Implikasinya Terhadap Responsivitas Bisnis (2024). https://doi.org/10.61132/jubid.v1i3.189
6. Dwivedi., R., Arvind, P.: AI robots in various sector. Int. J. Innov. Sci. Res. Technol. (2024). https://doi.org/10.38124/ijisrt/ijisrt24jun1345
7. Rachana, U.S., Vismay., Y, Raju, D.: Advanced robotics. Int. J. Adv. Res. Sci. Commun. Technol. (2024). https://doi.org/10.48175/ijarsct-19330
8. Santosh, R.D.: Influence of intelligent automation on industries and daily life. Advances in business information systems and analytics book series (2024). https://doi.org/10.4018/979-8-3693-3354-9.ch004
9. Maha, H., Mobeen, R., Nicholas, A.: Artificial intelligence and machine learning enhance robot decision-making adaptability and learning capabilities across various domains. Delet. J. (2024). https://doi.org/10.62304/ijse.v1i3.161
10. Akshay, B.R., Sini, R.P., Murugesh, T.S., Vasudevan, S.K.: Machine learning (2024). https://doi.org/10.1201/9781032676685

11. Mandeep, S., Subair, A, Liayakath, A.K.: Advances in autonomous robotics: integrating AI and machine learning for enhanced automation and control in industrial applications (2024). https://doi.org/10.61877/ijmrp.v2i4.135
12. Vaishali, J., Shiv, K.T.: Overview: machine learning (2024). https://doi.org/10.58532/nbennurch183
13. Manbir, S., Anuradha, S., Gurpreet, K.P.: Reinforcement learning: framework, applications and challenges. Int. J. Eng. Sci. Hum. (2024). https://doi.org/10.62904/s3qf5660
14. Sonal, A., Raj, G., Ashima, M.: Advancements in reinforcement learning. Int. J. Adv. Res. Sci. Commun. Technol. (2024). https://doi.org/10.48175/ijarsct-17820
15. Mandeep, S., Subair, A.L., Ali, K.: Advances in autonomous robotics: integrating ai and machine learning for enhanced automation and control in industrial applications (2024). https://doi.org/10.61877/ijmrp.v2i4.135
16. Ajay, V., Louise, A.D., Marija, S.: Reinforcement Learning and Machine ethics: a systematic review (2024). https://doi.org/10.48550/arxiv.2407.02425
17. Methods to predict the performance analysis of various machine learning algorithms (2022). https://doi.org/10.1201/9781003164265-3
18. Ritu, S., Mansi, M., Richa, G., Anuj, K., Richa, B., Ashish, G.: The performance analysis of health care automation using artificial intelligence model (2024). https://doi.org/10.1109/ic3se62002.2024.10593586
19. Nicolás, H., Francisco, A., Vicente, B.: Performance and energy efficiency: quantization of models for IoT devices (2023). https://doi.org/10.21203/rs.3.rs-3405705/v1
20. Georgios, S., Marinela, M., Maria-Evgenia, X., Nikos, P., Vasileios, T., Theodoros, B.: Unlocking the path towards automation of tiny machine learning for edge computing (2024). https://doi.org/10.1109/smartnets61466.2024.10577687
21. Emilian, V., Kais, G.: Supervised learning (2024). https://doi.org/10.1201/9781003254515-12
22. Supervised machine learning a brief survey of approaches. Al-Iraqia J. Sci. Eng. Res. (2024). https://doi.org/10.58564/ijser.2.4.2023.121
23. Keisuke, T., Lauren, T.: Supervised machine learning (2023). https://doi.org/10.1007/978-981-97-0217-6_8
24. Giuseppe, V.: Unsupervised learning (2024). https://doi.org/10.1201/9781003254515-11
25. Seghers, E.., Luis, A., Briceno-Mena, J., Romagnoli, A. (2023)
26. Maurizio, P.: Unsupervised machine learning methods. Springer textbooks in earth sciences, geography and environment (2023). https://doi.org/10.1007/978-3-031-35114-3_4
27. Tianjiao, D., et al.: Unsupervised manifold linearizing and clustering. https://doi.org/10.1109/iccv51070.2023.00502
28. Myky, T.: Unsupervised learning with restricted boltzmann machines and autoencoders (2023). https://doi.org/10.1007/978-1-4842-8931-0_5
29. Yamana, Y.: Deep learning and neural networks: methods (2023). https://doi.org/10.59646/csebookc7/004
30. Karel, H., Robert, S.: Deep learning concepts and datasets for image recognition: overview 2019 (2019). https://doi.org/10.1117/12.2539806
31. Sachin, S., Bhat., A.A., Venugopala, P.S.: Design and evolution of deep convolutional neural networks in image classification – a review. Int. J. Integrat. Eng. (2023). https://doi.org/10.30880/ijie.2023.15.01.019
32. Navdeep, S., Hiteshwari, S.: Convolutional neural networks-an extensive arena of deep learning. A comprehensive study. Arch. Comput. Meth. Eng. (2021). https://doi.org/10.1007/S11831-021-09551-4
33. Diego, M.-B., Luis, R., Luis, F.M.: RUMOR: reinforcement learning for understanding a model of the real world for navigation in dynamic environments (2024). https://doi.org/10.48550/arxiv.2404.16672

34. Sriram, S., Maggie, W.: RIDER: reinforcement-based inferred dynamics via emulating rehearsals for robot navigation in unstructured environments (2024). https://doi.org/10.1109/icra57147.2024.10611692
35. Tianci, G.: Enhancing robotic adaptability: integrating unsupervised trajectory segmentation and conditional ProMPs for dynamic learning environments (2024). https://doi.org/10.48550/arxiv.2404.19412
36. Noureldin, R., Mervat, M., Mahmoud, A.: Implementing deep reinforcement learning in autonomous control systems. J. Adv. Res. Appl. Sci. Eng. Technol. (2024). https://doi.org/10.37934/araset.41.1.168178
37. Basel, , Qiang, H.: Enhancing transfer learning reliability via block-wise fine-tuning (2023). https://doi.org/10.1109/icmla58977.2023.00064
38. Bhakti, G., Shinde, S., Kundan, S.: Impact of data visualization in data analysis to improve the efficiency of machine learning models (2024). https://doi.org/10.53555/jaz.v45is4.4161
39. Jacob, P.P., Davis, B., Cory, S., Jonathan, F.: Fast benchmarking of accuracy vs. training time with cyclic learning rates (2022). arXiv.org, https://doi.org/10.48550/arXiv.2206.00832
40. Pavel, K., Tomáš, F.: On scalability of predictive ensembles and tradeoff between their training time and accuracy (2017). https://doi.org/10.1007/978-3-319-70581-1_18
41. Matt, B., Nifesh, C., Hong-Linh, T., Krists, K., Kyle, C., Ian, F.: Measuring, quantifying, and predicting the cost-accuracy tradeoff (2019). https://doi.org/10.1109/BIGDATA47090.2019.9006370
42. Suyog, G., Wei, Z., Fei, W.: Model accuracy and runtime tradeoff in distributed deep learning: a systematic study (2017). https://doi.org/10.24963/IJCAI.2017/681
43. Mustafa, Baris Tradeoff Assessment of Deep Learning Models based on Accuracy, Time and Size
44. James, G.K.: Machine learning algorithms for predictive maintenance in manufacturing. J. Technol. Syst. (2024). https://doi.org/10.47941/jts.2144
45. Maki, K.H., Kamal, M.: Enhancing predictive maintenance hyperparameter optimization and adopted strategies (2024). https://doi.org/10.1109/icma61710.2024.10633049
46. Decision Trees, Random Forests and Boosting (2023). https://doi.org/10.1017/9781107588493.015
47. Hasan, A.S., Ali, K., Amani, S.: Random forest algorithm overview. Delet. J. (2024). https://doi.org/10.58496/bjml/2024/007
48. Giulia, D.T., Marta, M., Laura, P.: Unboxing Tree Ensembles for interpretability: a hierarchical visualization tool and a multivariate optimal re-built tree. EURO J. Comput. Optim. (2024). https://doi.org/10.1016/j.ejco.2024.100084
49. Khujaev, O., Nurmetova, B.B., Tohir, K.U.: Algorithms for selecting the most efficient method for solving classification problems (2023). https://doi.org/10.1109/apeie59731.2023.10347690
50. Jingjing, L., Jie-Peng, Y., Zhuo, W., Zhongyi, W., Lan, H.: Small sample time series classification based on data augmentation and semi-supervised learning. Inform. Technol. Control (2024). https://doi.org/10.5755/j01.itc.53.2.35797
51. Bo, Y., Kai, G., Tong, W., Min-Ling, Z.: Bridging the gap: learning pace synchronization for open-world semi-supervised learning (2024). https://doi.org/10.24963/ijcai.2024/593
52. Lareb, Z.K., João, P., Nelson, C., Andrea, S., Antonio, N., Nicola, S.: Model and data-centric machine learning algorithms to address data scarcity for failure identification. J. Opt. Commun. Network. (2024). https://doi.org/10.1364/jocn.511863
53. Lusiné, N., Martin, K., Ulf, L.: Benchmarking machine learning methods for the identification of mislabeled data (2024). https://doi.org/10.21203/rs.3.rs-4011683/v1
54. Lei, W., G., L.: Research on multi-robot collaborative operation in logistics and warehousing using A3C optimized YOLOv5-PPO model. Front. Neurorobot. (2024). https://doi.org/10.3389/fnbot.2023.1329589

55. Scalable multi-agent reinforcement learning for warehouse logistics with robotic and human co-workers(2022). https://doi.org/10.48550/arxiv.2212.11498

Enhanced Pedestrian Detection for Autonomous Vehicles Using Multi-localized Feature

Abhipsa Pattanaik[1], Amrapali Unkal[1], Isha Jagtap[1], D. Sangeetha[1(✉)], and S. R. Mugunthan[2]

[1] Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India
sangeetha.srit@gmail.com

[2] Department of Computer Science and Engineering, SRM Institute of Science and Technology (Ramapuram Campus), Chennai, India

Abstract. Pedestrian detection plays a vital task in numerous computer vision applications, particularly in autonomous driving systems. These systems heavily depend on the perception module, which must be both high-performing and efficient in order to make accurate decisions in real-time with low delay. Prioritizing the prevention of collisions with pedestrians is vital in every autonomous driving system. As a result, pedestrian detection is a fundamental component of the perception modules in these systems. Recent years have seen a rapid development of the pedestrian detection system, which aims to alleviate difficulties caused by changes in illumination, scale, appearance, blur, and occlusion. However, existing techniques does not support all the challenges simultaneously with better accuracy in real-time. The proposed human detection algorithm addresses these challenges by utilizing scale generation architecture, which fuses features such as the histogram of oriented gradients (HOG) with non-maximum suppression (NMS), local binary patterns (LBP), local ternary patterns (LTP), and the gaussian mixture model (GMM). The scale generation architecture model supports the detection of humans at various scales. The HOG model, in conjunction with NMS, LBP, and LTP, extracts gradient edge features and fine-grained texture features, which support the pedestrian detection in a variety of appearances and illuminations. GMM improves feature performance by providing more useful information. We feed these features into the support vector machine (SVM) with a radial basis function (RBF) kernel to detect the pedestrian. We validate the proposed pedestrian detection system on more challenging state-of-the-art databases, namely INRIA and Caltech, achieving an average accuracy of 97.05% with real-time support that takes 0.12s to detect the pedestrian.

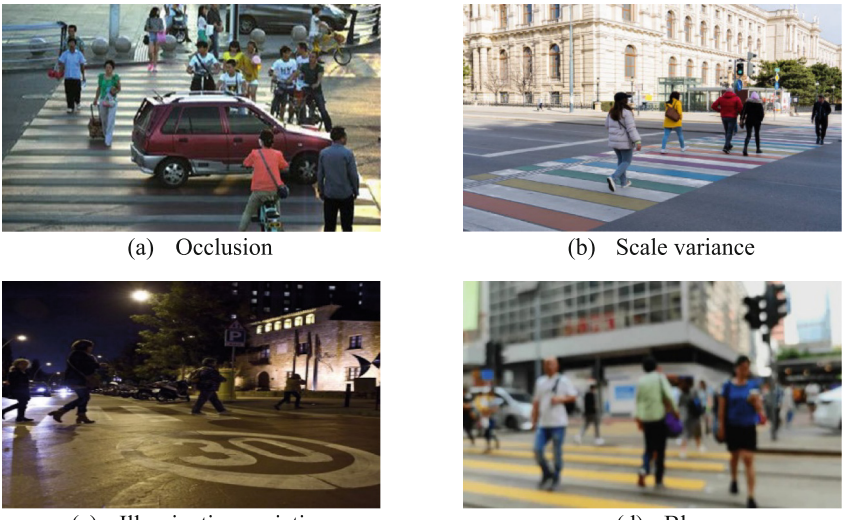
Keywords: HOG · NMS · LBP · LTP · pedestrian detection · autonomous driving system

1 Introduction

The intelligent transportation systems (autonomous driving), video surveillance, human-machine interaction and analysis has pushed computer vision to the forefront, particularly in terms of pedestrian detection [1]. To ensure, pedestrian vehicle safety implementation

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 101–113, 2026.
https://doi.org/10.1007/978-3-031-98138-8_8

of accurate pedestrian detection algorithm is crucial. Several approaches to pedestrian detection have been presented, ranging from classical feature-based approaches namely HOG [2], Haar – like features [3], Viola–Jones features [4], Texture [5], Local Binary Pattern (LBP) [6], integral channel features [7], scale invariant feature transform (SIFT) [8] to deep learning-based systems [9–13]. In classical feature-based methods, from an image the hand-crafted features are extracted that train the classifiers such as SVM, AdaBoost and random forest (RF) to detect the pedestrians by filtering the background. Whereas in the deep learning-based systems to find and classify the pedestrians in an image, the features are learnt automatically through the convolutional layers which is combination of multiple non-linear processes. The hand-crafted feature based are sufficient for simple scenarios. Nevertheless, the efficiency is poor and the performance is unsatisfactory. Breakthrough advancement in the field of deep learning, particularly the introduction of generic object detection, deep learning-based techniques for pedestrian detection have made significant advancements in terms of both speed and accuracy but require more resources since it is computationally extensive. Although existing pedestrian detection performance is still not quite as effective as human perception. The pedestrian detection systems still face many challenges such as large difference in appearance, scale variance, occlusion, illumination variations and blur as shown in Fig. 1.



(a) Occlusion (b) Scale variance

(c) Illumination variation (d) Blur

Fig. 1. Challenges in pedestrian detection: (a) Occlusion; (b) Scale variance; (c) Illumination variation; (d) Blur.

Researchers have been using the HOG as a feature-based technique to find pedestrians since 2005. The HOG is a feature descriptor that effectively captures gradient information in specific parts of an image, making it an excellent choice for pedestrian detection [2]. The gradient information, (i) captures the structural essence of pedestrians against a variety of backgrounds effectively, and (ii) is capable of detecting boundaries even

when the illumination changes. Parameters such as feature descriptor size and image dimensions determine the speed of pedestrian detection.

Most studies integrate HOG with the SVM classifier for pedestrian detection. The relationship between HOG features and linear SVM is crucial in pedestrian detection research, serving as a benchmark for detector performance. HOG features have high discriminatory power in image classification, leading to their widespread use in modern pedestrian detectors and other object detection tasks.

First, we will examine the theoretical basis of HOG. Subsequently, we will delve into its implementation and explore its various applications in pedestrian detection systems. The study examines significant advancements in the field, with a focus on real-time and tailoring the improvement of pedestrian detection accuracy and efficiency. The main aim is to detect the pedestrians for safety by integrating background modelling with GMM (Gaussian Mixture Model), feature extraction techniques, and SVM.

2 Related Work

Bahri et al. focus on reducing the high-dimensional feature vectors of the HOG using principal component analysis (PCA) in pedestrian detection. The PCA has applied to extracted HOG features, to reduce the dimensions from 3780 to 937 with increase in 1% of overall performance. SVM models with dimensionality-reduced features are experimented to find the suitable parameters and kernel for the dataset to attain high precision and recall rates. While PCA can improve computational efficiency, it may also result in a loss of discriminative information crucial for accurate pedestrian detection [14]. The combination of different feature descriptors, like Fourier descriptors and HOG, integrated with SVM has shown improved performance in real-time human detection on GPUs. The authors created a GMM model for extracting foreground images using CUDA. This paradigm optimizes thread-block configurations, reduces global memory transactions, maximizes shared memory utilization, and achieves high thread-block efficiency, resulting in decreased execution time [15] (Fig. 2).

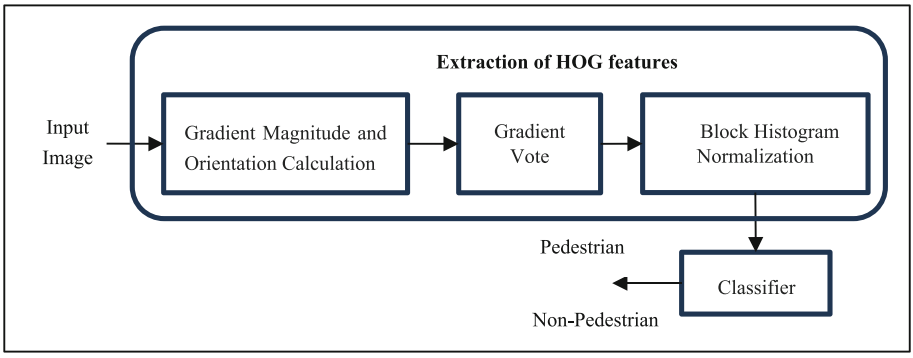


Fig. 2. HOG feature extraction based pedestrian detection

The incorporation of a mixture of Gaussian modelling enhances the robustness of the human detection system by effectively modelling complex background scenes. The background is represented as a mixture of Gaussian distributions, which can adapt varying lighting conditions, occlusions, and scene clutter, thereby improving overall detection performance [16]. Real-time pedestrian detection in videos is crucial for applications like surveillance in restricted areas. Traditional methods are computationally expensive, limiting their usefulness. Zhang et al. (2021) propose a two-step approach to address this challenge. A GMM separates moving objects from the background. To capture pedestrian shapes, the HOG features are extracted from these foreground regions, which are then trained using SVM classifier. It leads to reduced processing time, improved detection efficiency, and real-time applicability for video surveillance. However, the paper acknowledges limitations in handling complex backgrounds and occlusions and suggests further research to improve performance [17].

Traditional methods like HOG and SVM achieve excellent accuracy in controlled settings but struggle with real-world complexities that lead to missed detections, false alarms, and inaccurate bounding boxes. Tu et al. propose a method in which HOG is combined with SVM and face detection using Haar and AdaBoost. This approach leverages both body and facial information to improve detection accuracy and reduce false positives, particularly in complex scenes, especially regarding real-time suitability for resource-constrained devices [18]. Satyawan et al. investigate the pedestrian detection in autonomous vehicles using HOG features that operated in limited areas. The HOG features have proven effective in the unique challenges posed by limited operational environments, such as occlusions, background clutter, and varying pedestrian poses, which can significantly impact pedestrian detection accuracy in such scenarios. The authors emphasize the potential of HOG features and explores advancements to ensure reliable and safe pedestrian detection in challenging environments with the highest accuracy [19].

Angadi and Nandyal's research focus on applying HOG features with non-maximum suppression (NMS) for human identification in ATM surveillance systems. An NMS eliminates redundant bounding boxes, guaranteeing a single identification of each person in the presence of multiple detections. Using the well-known HOG-SVM, the adaptive background generation model, and the NMS together makes it possible to identify people very accurately in ATM video surveillance. When the NMS discards valid detections in scenarios with overlapping objects or closely spaced individuals, it has a negative impact on detection accuracy [20]. Zhou and Yu's research explores pedestrian detection using a combination of HOG and Local Ternary Pattern (LTP) features, leveraging the strengths of both for enhanced accuracy. The HOG features capture the shape and appearance information that is critical for pedestrian detection. However, the LTP features are known for their robustness to lighting variations. The authors train a SVM classifier with the combined feature vector. This classifier does a better job of detecting things and is more stable in challenging situations like changing lighting and occlusions [21].

Based on the proven success of HOG features in capturing pedestrian shapes, the work leverages K-means clustering to group similar HOG features due to its hardware-friendly nature and ability to avoid large training datasets. However, existing validity measures for K-means are found to be limited, prompting the introduction of the

novel normalization-based validity index (NbVI), a novel index designed for multi-dimensional HOG data. NbVI takes into account normalized inter- and intra-distances, making it more robust. Additionally, a diagonal selection strategy accelerates K-means convergence. Although FPGAs offer parallel processing capabilities, the overall system might still be computationally demanding, especially for high-resolution images or complex detection tasks. Finally, the entire system, including HOG feature extraction, NbVI validation, and adaptive K-means clustering is implemented on the hardware platform to achieve real-time performance in autonomous driving applications [22].

However, the existing approaches do not provide better performance in all the real-world situations at the same time. The proposed approach introduces a novel idea of scale generation and fusion of LBP, LTP, and HOG with NMS features, which achieves high detection accuracy under partial occlusion, scale variance, appearance variation, and illumination variation in less time.

3 Methodology

In the proposed design strategy, the preprocessing stages have been chosen to improve the algorithm's performance and address typical issues in pedestrian detection. This entailed numerous critical steps targeted at improving the dataset for algorithmic analysis. Initially, the images are resized to 128 x 64 pixels to ensure consistency and reduce computational complexity. RGB images are then transformed to grayscale, simplifying processing and increasing contrast for better feature detection. To reduce noise that could interfere with algorithm performance, a median filter with a 3 x 3 kernel is applied to each image. Additionally, the data augmentation techniques namely horizontal flipping with a 50% chance during training increases the diversity of training data that improves the capacity of the model to generalize the previously unknown circumstances. Finally, the pre-processed images were saved into a new directory structure that mirrored the original dataset, with distinct folders for training and testing sets, each containing positive and negative samples. Collectively, these preprocessing techniques enhance the robustness.

Multi-scale human detection requires multiple classifiers, which complicates the training phase. To get around this problem, the proposed method uses a scale generation block. This block generates an input image at various scaled versions during the training stage. The scale factor, which is the ratio between successively generated dimensions of an image, influences pedestrian detection accuracy. The input image is generated at nine scales to give better detection results by using a trial-and-error approach.

Feature extraction is carried out on scaled version of the images utilizing HOG with NMS, LBP, and LTP methods. These approaches are used to extract different patterns and textures from image, allowing for accurate pedestrian detection. The HOG with NMS approach measures gradient orientations in specific image regions, resulting in a detailed depiction of object structure and appearance. LBP encode texture patterns using binary comparisons of pixel intensities within local picture patches, ensuring resistance to changes in illumination and noise. Furthermore, LTP build on LBP by taking into account both binary and ternary patterns in image textures, capturing finer details and changes in texture patterns. The extracted features are augmented to enhance the robustness against the appearance variations, noise also improves the performance on the minority classes

if there is a class imbalance in the dataset. The augmented data is normalized to ensure that the augmented data maintains the same properties as the original data, leading to more stable and reliable model training and evaluation. Extracting the GMM parameters of the normalized data helps to gain a deeper understanding underlying structure of the data. The normalized data and GMM parameters train the SVM classifier for detecting the pedestrian. Figure 3 shows the proposed feature extraction algorithm block diagram for pedestrian detection system.

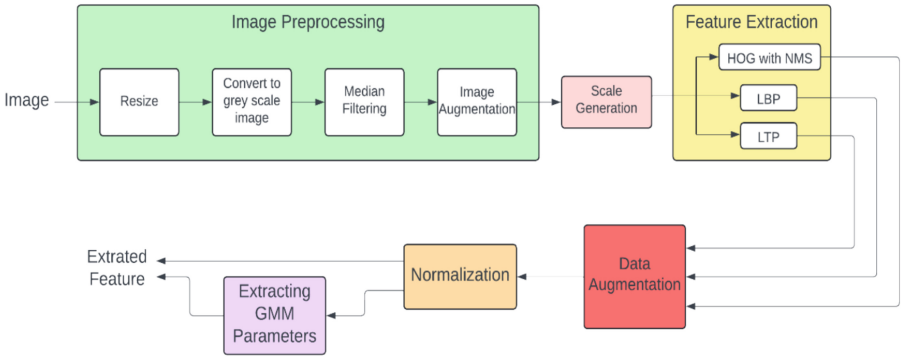


Fig. 3. Pedestrian detection system: proposed feature extraction algorithm

The HOG feature descriptor is useful for pedestrian detection because it captures the human silhouettes using the gradient pattern. This method divides the image into small linked sections known as cells, and for each cell, a histogram is plotted for directions of the gradient [23]. The combination of these histograms yields the feature description. Key steps include computation of gradient magnitude and orientation, NMS, gradient vote and block histogram normalization. For each pixel f (x, y), the gradients in x- and y- direction is calculated using Eq. 1(a) and (b) respectively,

$$G_x(x, y) = f(x + 1, y) - f(x - 1, y) \tag{1(a)}$$

$$G_y(x, y) = f(x, y + 1) - f(x, y - 1) \tag{1(b)}$$

The gradient magnitude and orientation are calculated based on the gradients using Eqs. (2), (3) and are given by,

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{2}$$

$$\theta_G(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \tag{3}$$

NMS processes the gradient magnitude. In NMS, the gradient magnitude of the corresponding neighbours is extracted where the orientation of the current pixel matches

with the orientation of eight neighbours. Next, we compare the extracted gradient magnitude with the current pixel magnitude, retaining it if it has the highest value; otherwise, we suppress it. We process all the pixels in the image in a same way.

The gradient orientation is divided into nine angular bins and are weighted using the corresponding non-maximum suppressed gradient magnitude to avoid aliasing effect. The four cells are combined to form the block. The overlapping blocks of cells are employed, and the histogram of gradients in each block is normalized. The normalizing stage in HOG processing decreases the influence of lighting differences across images, making it more trustworthy under varying lighting conditions.

3.1 LBP

LBP tags an image pixel by thresholding the neighborhood of each pixel and converts that into binary integer and is an efficient texture operator. Texture encoding compares each pixel intensity to that of its eight neighbours in the clockwise direction and the comparison yields a binary number, given using Eq. (4).

$$LBP(x, y) = \sum_{p=0}^{7} s\big(g_p - g_c\big) \times 2^p \tag{4}$$

where g_c is the centre pixel's intensity, g_p is the neighbouring pixel's intensity, and $s(x)$ is 1 if $x \geq 0$ else 0.

Histograms of the resultant LBP values are generated for sections of the picture to produce a feature descriptor. LBP excels in detecting fine-grained texture patterns, which aids in distinguishing pedestrian apparel textures from the backdrop.

3.2 LTP

LTP expand LBP to three values, boosting noise resistance by recording each pixel's intensity in three ways (below, equal, or above the centre pixel intensity). For pedestrian detection, similar to the LBP, LTP has three-level encoding and is given using Eq. (5),

$$LTP(x, y) = \sum_{p=0}^{7} t\big(g_p - g_c\big) \times 3^p \tag{5}$$

where $t(x) = 1$, if x is above a threshold, 0 if it is within the threshold range or else -1 otherwise. Separate histograms for zero, positive and negative values are computed by concatenating them. LTP improves LBP's robustness against noise by separating pixel differences into three levels, allowing for more tolerance against noise and minor fluctuations in pixel values. By employing three values rather than two (as in LBP), LTP may extract more detailed and discriminative characteristics from the texture, improving detection accuracy. In LTP, the thresholding may be adjusted to maximize performance in response to specific environmental variables such as illumination or background complexity.

3.3 Augmentation of the Extracted Features

We augmented the extracted features, complementing the augmentation we performed during preprocessing. This strategy has substantial advantages for boosting model performance. For starters, enhancing features directly increases the variety of the feature space, which improves the model's resilience and generalizability. By incorporating changes in the feature vectors, the model is better able to handle the varied circumstances and pedestrian appearance. Furthermore, augmentation on the extracted features addresses class imbalance concerns by providing synthetic examples of minority classes, resulting in a more balanced dataset and reducing biases in model training. Furthermore, enhancing features with random noise or transformations strengthens the model's robustness to fluctuations in input data, resulting in improved performance on previously unknown test data. In general, extracted features improves the model performance because it balances the data, and makes the model better at adapting to different pedestrian situations that happen in real life [24]. After augmenting the extracted features, we used normalization to ensure that all feature dimensions were on the same scale. This stage is critical for avoiding features with higher magnitudes from dominating the learning process, resulting in a more balanced and stable training approach.

3.4 Fitting a GMM

GMM is important in pedestrian identification because it compresses the underlying feature distribution into a combination of Gaussian components, allowing vital statistical features to be extracted from data [25]. Fitting a GMM entailed applying it to normalized data in order to capture the underlying feature distribution and give a compressed form. Parameters like as means, variances, and covariances were retrieved as additional features, hence improving the feature representation. These GMM-based features were subsequently combined with the original features to form a fused representation that included both data characteristics and statistical traits. This extensive feature set allowed the model to make more informed judgments during training and inference, possibly increasing its discriminative strength and capacity to capture complicated patterns. GMM improves feature representation by capturing complex patterns and variations found in pedestrian images using parameters such as means, variances, and covariances. Compressed feature space retains critical statistical information, allowing for efficient storage and processing in pedestrian detection systems.

3.5 Support Vector Machine (SVM) with Radial Basis Function (RBF) Kernel

In pedestrian detection, the SVM with RBF kernel helps in learning decision boundaries in high-dimensional feature spaces, thereby enabling effective categorization of pedestrians [26]. We then train a SVM classifier with an RBF kernel using the processed data from the training set. The pedestrian detection challenge benefits greatly from the RBF kernel's recognition of complex, nonlinear connections in data. SVM classifier successfully learn decision boundaries in high-dimensional feature spaces, allowing for correct classification even in scenarios with complex data distributions. Furthermore, the RBF

kernel implicitly transferred the features into a higher-dimensional space, allowing the SVM to better capture complex patterns and changes in the data.

When the linear separation is insufficient, we use the RBF kernel's non-linear mapping to identify pedestrians in a variety of environments with complex backgrounds and lighting conditions. The model became more stable and generalizable due to the RBF kernel's ability to show complex interactions between features, enabling it to perform better on previously unexplored data examples. Adding the RBF kernel to the SVM architecture made it easier to capture the complex interactions that happen in pedestrian detection tasks. This led to better model accuracy and performance.

4 Experimental Result and Discussion

We used the INRIA [27] and Caltech [28] datasets to validate the proposed pedestrian detection algorithm. We pre-processed the dataset to enhance its appropriateness for pedestrian detection applications. This dataset includes images with varied resolutions, various postures, orientations, scales, lighting conditions, complex backgrounds, and occlusion. The classifier trains on a set of 1239, and 1218, and tests on 563 and 453 positive and negative images respectively from the INRIA dataset. We extract training and test samples every 30^{th} frame from the Caltech dataset. The proposed algorithm works well with varying pose and scale as shown in Fig. 4.

Furthermore, we compare the detected results to ground truth images to assess pedestrian detection's accuracy. If the detected result matches the ground truth images, it is accurate. Table 1 compares the accuracy rates of the proposed and conventional HOG algorithms. The investigational results show the effectiveness of the proposed pedestrian detection technique, as it achieves an overall detection accuracy of 97.05%. Additionally, it exhibits a low false and missed detection rate of 0.02% and 0.15%, respectively. The results demonstrate superior accuracy and computational efficiency as compared to existing methods. Preprocessing techniques, such as image scaling, grayscale conversion, and data augmentation, significantly improved the performance of the algorithm by enhancing feature representation and reducing noise.

The total detection accuracy has enhanced by the effective collection of pedestrian features using feature extraction techniques, including HOG with NMS, LBP, and LTP. Robust and reliable pedestrian recognition was made possible by the combination of the GMM for feature compression and the SVM with RBF kernel for decision boundary learning. Overall, the proposed approach shows promise for practical pedestrian detection applications and has room for improvement and additional optimization. The proposed pedestrian detection algorithm takes 0.12s to detect the pedestrian in the single image and is much suitable for the real-time applications. Our method achieves a balance between speed and accurateness in the detection approach, as it achieves an overall detection accuracy of 97.05%. Additionally, it exhibits a low false and missed detection rate of 0.02% and 0.15%, respectively. The results demonstrate superior accuracy and computational efficiency as compared to the existing methods. Preprocessing techniques, like image scaling, grayscale conversion, and data augmentation, significantly improved the performance of the algorithm by enhancing feature representation and reducing noise.

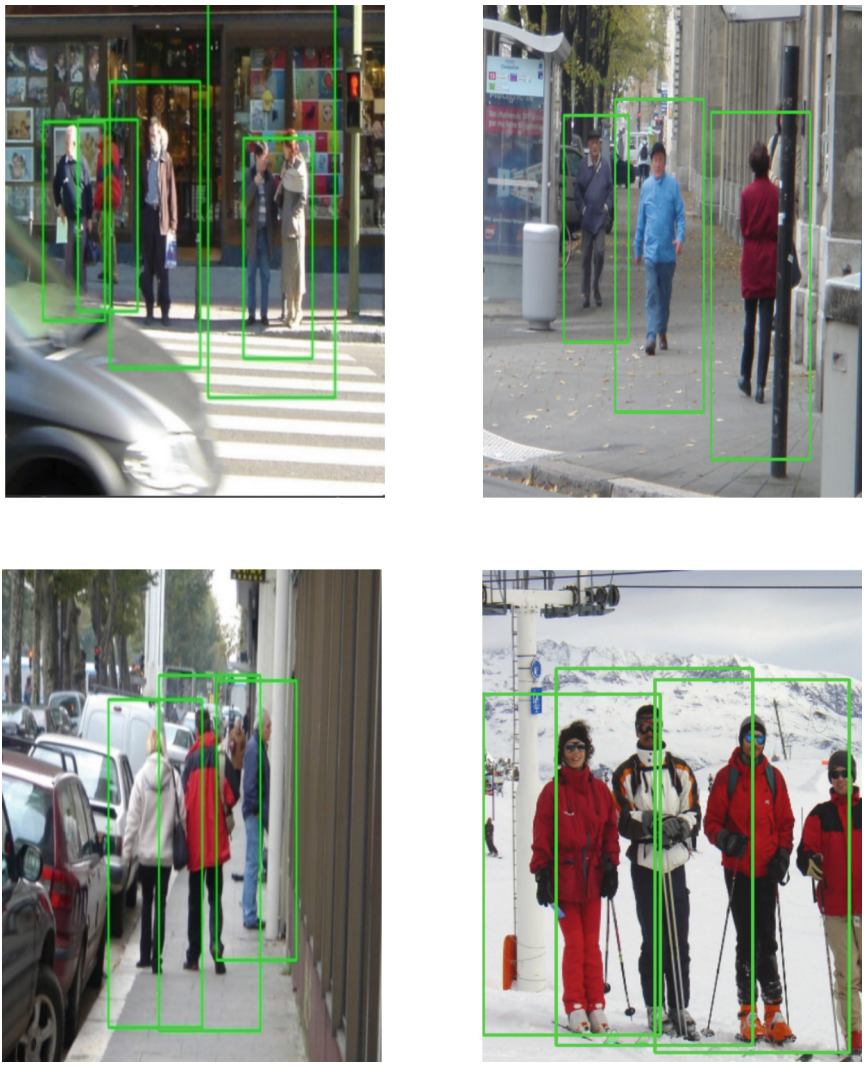


Fig. 4. Pedestrian Detection on INRIA and Caltech dataset.

The total detection accuracy has enhanced by the effective collection of pedestrian features using feature extraction techniques, including HOG with NMS, LBP, and LTP. Robust and reliable pedestrian recognition was made possible by the combination of the GMM for feature compression and the SVM with RBF kernel for decision boundary learning. Overall, the proposed approach shows promise for practical pedestrian detection applications and has room for improvement and additional optimization. The proposed pedestrian detection algorithm takes 0.12s to detect the pedestrian in the single image and is much suitable for the real-time applications. Our method achieves a good balance between speed and accuracy.

Table 1. Comparison of accuracy rate

		INRIA [27]	Caltech [28]
Conventional HOG	Detection Results	TP: 543, TN: 430, FP: 23, FN: 20	TP: 494, TN: 840, FP:225, FN: 475
	Accuracy	95.76% (973/1016)	65.58% (1334/2034)
Proposed HOG	Detection Results	TP: 555, TN: 450, FP: 3, FN: 8	TP: 921, TN: 1015, FP: 50, FN: 48
	Accuracy	98.92% (1005/1016)	95.18% (1936/2034)

5 Conclusion

The research work presents a novel pedestrian detection algorithm that fuses the HOG with NMS, LBP, LTP, and GMM features. The extracted features perform well under varying appearance, illumination, and scale. The extracted features train an SVM with an RBF kernel classifier to detect pedestrians. The novel pedestrian detection algorithm performs better because it achieves detection accuracy of 98.92% and 95.18% for the INRIA and Caltech datasets, respectively, with significantly reduced detection time. Furthermore, the proposed algorithm achieves a miss and false detection rate of 0.15%, 0.02%, respectively which are low. However, we identified issues such as computational complexity and overfitting, suggesting potential avenues for future research.

Disclosure of Interests. The authors have no competing interests.

References

1. Gohar, A., Nencioni, G.: The role of 5G technologies in a smart city: the case for intelligent transportation system. Sustainability **13**(9), 5188 (2021)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 886–893. IEEE (2005)
3. Oren, M., Constantine P., Pawan S., Edgar O., Tomaso P.: Pedestrian detection using wavelet templates. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 193–199. IEEE (1997)
4. Viola, S.: Detecting pedestrians using patterns of motion and appearance. In: Proceedings Ninth IEEE International Conference on Computer Vision, pp. 734–741. IEEE (2003)
5. Gavrila, D.M., Munder, S.: Multi-cue pedestrian detection and tracking from a moving vehicle. Int. J. Comput. Vision **73**(1), 41–59 (2007)
6. Wang, X., Han, T. X., Yan, S.: An HOG-LBP human detector with partial occlusion handling. In: 12th International Conference on Computer Vision, pp. 32–39. IEEE (2009)
7. Dollar, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: Proceedings of the British Machine Vision Conference, London (2009)

8. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh International Conference on Computer Vision, pp. 1150–1157. IEEE (1999)
9. Li, C., Xinggang, W., Wenyu, L.: Neural features for pedestrian detection. Neurocomputing **238**, 420–432 (2017)
10. Szarvas, M., Yoshizawa, A., Yamamoto, M., Ogata, J.: Pedestrian detection with convolutional neural networks. In: IEEE Proceedings. Intelligent Vehicles Symposium, pp. 224–229. IEEE (2005)
11. Zhang, L., Lin, L., Liang, X., He, K.: Is faster R-CNN doing well for pedestrian detection? In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. ECCV 2016. LNCS, vol 9906. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_28
12. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Ann. Stat. **28**(2), 337–407 (2000)
13. Appel, R., Thomas, F., Piotr, D., Pietro, P.: Quickly boosting decision trees–pruning underachieving features early. In: International Conference on Machine Learning, pp. 594–602. PMLR (2013)
14. Putri, S. A., Hasibuan, Z. A., Purwanto, Soeleman, M.A.: Dimensional reduction with PCA for feature selection in pedestrian detection. In: 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 340–347 (2021)
15. Bahri, H., Chouchene, M., Sayadi, F.E., Atri, M.: Real-time moving human detection using HOG and Fourier descriptor based on CUDA implementation. J. Real-Time Image Proc. **17**, 1841–1856 (2020)
16. Akila K., Pavithra, P.: Optimized scale invariant HOG descriptors for object and human detection. In: IOP Conference Series: Materials Science and Engineering, vol. 1119(1), pp. 012002 (2021)
17. Zhang, Y., Guo, K., Guo, W., Zhang, J., Li, Y.: Pedestrian crossing detection based on HOG and SVM. J. Cyber Secur. **3**(2), 79–88 (2021)
18. Tu, R., Zhu, Z., Bai, Y.: Improved pedestrian detection algorithm based on HOG and SVM. J. Comput. **31**(4), 211–222 (2020)
19. Satyawan, A.S., Fuady, S., Mitayani, A., Sari, Y.W.: HOG based pedestrian detection system for autonomous vehicle operated in limited Area. In: 2021 International Conference on Radar. Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Bandung, pp. 147–152. IEEE, Indonesia (2021)
20. Angadi S., Nandyal, S.: Human identification using histogram of oriented gradients (HOG) and non-maximum suppression (NMS) for ATM video surveillance. Int. J. Innov. Res. Comput. Sci. Technol. 2347–5552 (2021)
21. Zhou, H., Yu, G.: Research on pedestrian detection technology based on the SVM classifier trained by HOG and LTP features. Futur. Gener. Comput. Syst. **125**, 604–615 (2021)
22. Li, T., Yitao, M., Hui, S., Tetsuo, E.: FPGA implementation of real-time pedestrian detection using normalization-based validation of adaptive features clustering. IEEE Trans. Veh. Technol. **69**(9), 9330–9341 (2020)
23. Zhao, Y., Zhang, Y., Cheng, R., Wei, D., Li, G.: An enhanced histogram of oriented gradients for pedestrian detection. IEEE Intell. Transp. Syst. Mag. **7**(3), 29–38 (2015)
24. Alonso, I.P., et al.: Combination of feature extraction methods for SVM pedestrian detection. IEEE Trans. Intell. Transp. Syst. **8**(2), 292–307 (2007)
25. Zhang, M., Jin, J.S., Wang, M., Tang, B., Zheng, Y.: Pedestrian intrusion detection based on improved GMM and SVM. In: Thirteenth International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 311–315 (2016)
26. Cao, H., Naito, T., Ninomiya, Y.: Approximate RBF kernel SVM and Its applications in pedestrian classification. In: The First International Workshop on Machine Learning for Vision-based Motion Analysis - MLVMA'08, Marseille, France (2008)

27. INRIA person dataset. pascal.inrialpes.fr/data/human/
28. Caltech pedestrian dataset. http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

Enhancing Lead Score Conversion Rate Using Logistic Regression

Shirish Joshi(✉)

Symbiosis Institute of Computer Studies and Research, SIU, Pune, India
shirishjoshi2911@gmail.com

Abstract. Businesses frequently collect massive amounts of information like surfing behavior, email activity, and other personal data. By applying statistical analysis to estimate a contact's transaction likelihood, this data can provide a significant competitive edge. Most businesses struggle with weeding out potential consumers from their audience. These consumers called leads are very beneficial for various purposes like enhancing a firm's sales or expanding the business etc. The purpose of conducting this work is to find potential leads from existing consumers so that a company named X Education can enhance its conversion rate. Machine learning techniques are used for this work. We applied a logistic regression model to the data to convert the poor leads into hot leads. A generalized linear model was also used for getting the potential information of a lead and working on them. After applying the regression model the conversion rate was found to be increased to 70%.

Keywords: Score Conversion · Logistic Regression · Machine Learning · Linear Model

1 Introduction

1.1 Background

Business development, also known as customer acquisition, is a vital component of any firm that provides goods or services, regardless of how many of the leads originate from potential leads or current clients. Certain leads will always outperform others in terms of conversion to bookings. Lead scoring is a mechanism for ranking your leads based on their likelihood of converting into customers. Znbound was among the first HubSpot Partners to deliver Lead Scoring to its clients. To put it another way, we assign a score to each action a lead does on our website, app, or elsewhere and utilize it to rank our leads.

A lead is defined as "a documented stated interest in the company's goods or services," regardless of whether the interest is from a new prospect or an existing client. A lead characterization theory's essential is determining which independent observable lead qualities have a major influence on the likelihood of lead conversion to a booking, and why.

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 114–130, 2026.
https://doi.org/10.1007/978-3-031-98138-8_9

Although there is a substantial body of research on consumer purchasing behavior, there is a surprising lack of literature on the categorization of sales leads, and no clear unifying theory has evolved. Additionally, corroborating evidence is scarce. Nonetheless, the literature on buying behavior does give some relevant information on the "determinants of client purchase decisions," which may be beneficial for lead conversion modeling. First, the large research on consumer purchasing behavior is examined, followed by a discussion of the considerably more restricted material on lead characterization.

1.2 Need

Most organizations run marketing efforts to reach out to consumers and produce leads for the business, but the difficult challenge here is to categorize the leads so that potential prospects may be recognized and converted to customers. Finding promising leads and focusing on them can save the organization a lot of time and money, which can be invested in other business efforts.

1.3 Problem Definition

Most organizations run marketing efforts to reach out to consumers and produce leads for the business, but the difficult challenge here is to categorize the leads so that potential prospects may be recognized and converted to customers. Finding promising leads or hot leads and focusing on them can save the organization a lot of time and money, which can be invested in other business efforts.

1.4 Objectives

The main issue that companies need to consider is the poor success rate of leads on the customer side. The primary goal of this research is to build and construct a model in which a lead value is allocated to each lead so that certain clients with a larger lead value own a greater conversion probability and clients with a negative outcome value have a low occupancy chance. The objective of this work is to look towards the most significant potential, or those clients who are most probable to turn into paying clients.

2 Literature Review

Etminan, A. (2021) The study focused on creating a lead score model for businesses to qualify leads based on three factors: explicit parameters, implicit parameters, and negative parameters. These metrics were chosen because they indicate the lead's behavior and level of involvement. By giving scores to each parameter, the total lead score was determined. Each parameter's value was decided based on its relevance and in cooperation with the sales team. The leads were then classed as hot, warm, or cold depending on the lead score, and a lead score matrix based on explicit and implicit scores was constructed. The research's main result was that from a collection of 1900 leads, 21% are hot leads that are revenue eligible, 35% are warm leads that are promotionally trained, and 44% are cold leads. Leads were qualified using the lead score formula. The matrix results

suggested that 60% of the leads were acceptable. As a result, it assisted digital enterprises in filtering out unqualified leads and managing leads more effectively, improving the quality of leads given to clients. This increased the customer goal ratio.

Benhaddou & Leray (2018) In this research, authors showed how to use a Bayesian network to develop a Lead scoring model with a limited quantity of data. Bayesian classifiers were conceptual modeling frameworks that may be formed using domain expertise, furthermore, their capacity to handle ambiguity. In this unique scenario, authors recommended that they design a Lead rating engine on skills and knowledge and use standard algorithms to simplify the elements of the model. They offered three approaches to calculating the variables of our NoisyOr scenarios. The technique was validated using the only data provided, with high precision and recall outcomes on a limited collection of 23 cases.

Sales leads are the backbone of modern industries, but deciding those leads are most probable to lead to bookings sometimes relies on instinct or assumption. As an outcome, assets are wasted, sales predictions are wrong, and revenues may be lost. A supply chain cost that can estimate whether any leads will convert according to details included in the leads individually would be extremely useful. Many lead characterization models have been developed; however, none has been field tested or is based on good theory. Monat (2011) This paper developed a qualitative model based on the good conceptual framework of business purchase choice factors. It was steady with certain of the prior unauthenticated theories, but it was more complete and hence should be more precise. The model identifies 16 distinct lead indicators that were often visible and should be gathered and analyzed for each lead to assist estimate its likelihood of translation to a booking. A preliminary qualitative time required of the strategy revealed that the model predicted how each lead will indeed transform into revenues and the ones that would not.

Diogo et al. (2020) The objective of this endeavor was to create a way to resolve the issue of sales leads as a vital phase in consumer implementation. However, this effort was aimed at a particular enterprise, it was a problem and a worry for many businesses, whatever their sector or size. Considering the challenge, two ideas were put forward: a conventional lead assessment instrument with the innovation of a lead leaderboard as a supplement to an agile approach, and the use of data analytics and ml algorithms to establish a prediction model that can forecast the probability of a lead becoming a customer, i.e., whether it will be labeled as Won or Lost. Several clustering algorithms were used, verified by the conclusions of articles done by the science establishment: Decision Tree, Random Forest, Gradient Boosted Trees, and Logistic Regression. Various circumstances had been created and assessed in terms of the number of entries on the collection with null values and the imbalance of facts across the distinct types: Won and Lost. To solve either one, several sampling approaches (under-sampling, oversampling, and SMOTE) were used to tackle the issue of unequal distribution of classes. The decision tree technique in Case 2 was just the highest-rated performing approach among the four situations.

Wu et al. (2023) The purpose of this research was to explore and assess the published studies on lead rating algorithms and their influence on profitable growth. To investigate lead ranking algorithms, a thorough study of the literature was performed. 44 papers

satisfied the parameters and became part of the assessment. To assess the influence of lead generation techniques on profitable growth, 14 measures were identified. Forecasting lead ranking approaches were projected to require new lead assessment approaches as the usage of information extraction and machine intelligence techniques increased in the new technologies, as they favorably affect sales success. Comparatively substantial expense of building and sustaining forecasting lead generation systems, it is still advantageous to replace conventional lead scoring models due to their greater effectiveness and efficiency. According to this survey, classification was the most often used data mining paradigm, whereas decision tree and regression analysis were the most widely used algorithms across all forecasting lead management models. This work contributed by systematizing and proposing whether a machine learning approach should be utilized to develop predictive lead score projections based on the authenticity of various types of data inputs. Furthermore, this work provided both academic and applied research prospects about lead scoring.

Several companies fail with weeding out target users from existing consumers. Etminan (2021) This project presented a mechanism for classifying prospective buyers from arbitrary online searches using user details. The framework relied on Forecasting Lead Ranking, which divides clients into groups depending on their chance of acquiring a product. The strategy, on the other hand, seeks to forecast user transformation, or if a consumer had the chance to turn into a client. The sorting challenge was completed using six supervised machine-learning techniques. Several quantization methods were used on the sample data to adjust for the large disparity in the raw data. The training algorithm was a mix of classification and subsampling approaches with the largest mean accuracy score. Furthermore, by assessing several components ordering and balancing techniques, this work aimed to measure the influence of characteristic values. Many sets of values were generated using the techniques and assessed by retraining a KNN classifier upon those weighted features. The difference in average accuracy acquired from the basic KNN was used to assess the effectiveness of sorting and weighting methods.

Forecasting lead rating is a popular study area for companies seeking novel prospects. Jadli et al. (2022) The studies examined ways to employ algorithms developed with machine learning to streamline the procedure of candidate discovery in aiming to supplant the old ranking method in this study. Various ML models, such as Logistic Regression (LR) and Random Forest (RF), were utilized to categorize prospective leads as approved or unskilled. The two highest methods (RF and DT) together had accomplished good results, and they could forecast whether consumers will submit a request on the company's site based on past and behavioral data with a consistency of 93.02% and 91.47% and a consistency of 92.25% and 88.37%, accordingly.

Indexed (2020) The study gave a reduced edition of their suggested lead qualifying process for creating a lead generation program for tiny businesses. It evaluated different factors for calculating lead ranks and used a simple chart to reduce the model's difficulty. The characteristics, such as firm size, signups, email id, and unsubscribe, are critical and had a significant impact on lead scores. A solid lead point total system provided a path for the firm to boost sales efficiency and profitability. As a result, by emphasizing unfavorable variables in the model, the constraints of traditional lead score models might be addressed. The mean lead rank of C-level (12.53) was greater than operations (9.07),

which was greater than Manager (6.81), which was more than E-commerce (5.51). It was also discovered that such classification categories comprised c-level and operations were the firm's ideal and useful potential customers. This knowledge will be beneficial throughout the first step of the lead transfer process, which is lead creation. As a result, it aids the group's forecasting and channel insight. Based on the findings, it was also proposed that targeted promotions be issued upon each set of classification types to engage folks with the help of the Potential User Profile.

Internet advertising has grown into an integral aspect of how businesses acquire new clients. Digitalization has caused significant shifts in how buyers and sellers seek knowledge and do analysis before purchasing anything. Throughout the B2B segment, there has been a big difference in how online technology impacts purchase decisions. Lindahl et al. (2017) The objective of this study was to look at a lead rating through the lens of a B2B sales funnel and see how to lead rating may help Visma Benefit have a more successful and cost-effective sales strategy. The goal was to create a proposal for an ideal lead rating system for Visma Benefit, as well as which sales promotions are appropriate for the various lead significance levels. The analysis reveals that there exist numerous approaches to making the best utilization of a lead ranking model for Visma Benefit. In addition, the outcome included a finished lead participation assessment and lead description evaluation, as well as a lead score structure that sets the criteria for each lead, mean scores. Furthermore, the output included applicable marketing activities for every lead mean score in the model.

Duncan & Elkan (2015) This work demonstrated what to do to train statistical algorithms to understand how revenue leads progress from the earliest moments to final achievement or defeat. The authors provided two concepts, DQM (direct qualification model) and FFM (full funnel model), that might be used to grade preliminary leads according to their likelihood of transformation to a potential lead, the likelihood of a possible sale, and/or predicted income. The massive number of records generated by consumer interaction planning or commercial machine intelligence was used in training. Traditional lead scoring methods, which were custom and hence failure and non-probabilistic, can be replaced with a classification model. DQM and FFM were intended to eliminate selection bias induced by accessible data that was based on a standard lead score method. The experimental findings were displayed using real-world sales data from two firms. The training data contained demographic and behavioral details for every prospect. Both techniques obtained strong AUC ratings for both firms. They brought in a 307% spike in the amount of sales performance and a significant increase in overall revenue for one organization. In addition, we discussed the DQM method's real outcomes. These findings indicated that the strategy had other advantages, such as requiring less time to qualify leads and requiring fewer phone calls to organize a product demo.

Nygård & Mezei (2020) Scientific work was described in this piece to determine the viability and effectiveness of using several models based on machine learning to automate lead scoring as a substitute to the current extensively utilized human lead scoring procedure. The authors examined the most extensively used machine learning algorithms from the research in this post. In addition, as a follow directly, they recognized numerous plausible grouping methods for finding necessary action for prospects

that did not result in an itemized invoice within the period of the data processing, and they assessed these methods in the context of different classifiers and biases introduced during the simulation analysis. Authors discovered that, while gathering and analyzing specific records on prospective leads is difficult, a high-performance comparison might be obtained even when correcting for model construction bias. Furthermore, they discovered that the arbitrary forest approach outperformed all other models in terms of its general efficiency.

Effective advertising organizations in both commercial and retail marketplaces typically spend a significant amount of their marketing expenditures on customer acquisition via advertising consultants and transformation by salespersons. Yet, because of the complex channel identification challenge, such an advertising interface design has frequently been determined to be inefficient. Banerjee & Bhardwaj (2019) employed analytical models to develop effective sales compensation solutions to overcome the multi-channel attribution challenge. Contracts incorporating revenue benefits, lead qualifying, and sales autocracy, according to the findings, created a gap between the best and the achieved profit owing to budget balance, lead qualification expenses, and the sales force's lack of marketing specialty, respectively. Increased risk avoidance favors sales exclusivity and lead qualifying deals over profit strategy will enable, whereas increased overall ambiguity supports lead screening. When ambiguity is low, a specific sort of challenge (or stack ranking-based compensation) yields the first-best profit.

3 Methodology

3.1 Process of Lead Conversion

Fig. 1. Lead Conversion Process

In the above Fig. 1 you recognize, there have been many leads created in the starting point (top), but just a couple of them become paying clients in the second step. To acquire a greater rate of conversion in the middle stage, you must develop possible leads properly which includes giving needful information to the leads, collaborating with them, etc.

3.2 Dataset

For this project, a lead sample was used from the previous era with almost 9000 sample points. This information contains numerous parameters including Lead Number,

Converted, Lead Source, Lead Profile, Total Time Spent on Website, Total Visits, Last Activity, and so on that may or may not be relevant in determining whether a lead will be transformed. In this case, the most useful feature is the column 'Converted,' which indicates whether a previous lead was transformed or not, with 1 indicating that it was converted and 0 indicating that it was not converted.

#	Column	Non-Null Count	Dtype
0	Prospect ID	9240 non-null	object
1	Lead Number	9240 non-null	int64
2	Lead Origin	9240 non-null	object
3	Lead Source	9204 non-null	object
4	Do Not Email	9240 non-null	object
5	Do Not Call	9240 non-null	object
6	Converted	9240 non-null	int64
7	TotalVisits	9103 non-null	float64
8	Total Time Spent on Website	9240 non-null	int64
9	Page Views Per Visit	9103 non-null	float64
10	Last Activity	9137 non-null	object
11	Country	6779 non-null	object
12	Specialization	7802 non-null	object
13	How did you hear about X Education	7033 non-null	object
14	What is your current occupation	6550 non-null	object
15	What matters most to you in choosing a course	6531 non-null	object
16	Search	9240 non-null	object
17	Magazine	9240 non-null	object
18	Newspaper Article	9240 non-null	object
19	X Education Forums	9240 non-null	object
20	Newspaper	9240 non-null	object
21	Digital Advertisement	9240 non-null	object
22	Through Recommendations	9240 non-null	object
23	Receive More Updates About Our Courses	9240 non-null	object
24	Tags	5887 non-null	object
25	Lead Quality	4473 non-null	object
26	Update me on Supply Chain Content	9240 non-null	object
27	Get updates on DM Content	9240 non-null	object
28	Lead Profile	6531 non-null	object
29	City	7820 non-null	object
30	Asymmetrique Activity Index	5022 non-null	object
31	Asymmetrique Profile Index	5022 non-null	object
32	Asymmetrique Activity Score	5022 non-null	float64
33	Asymmetrique Profile Score	5022 non-null	float64
34	I agree to pay the amount through cheque	9240 non-null	object
35	A free copy of Mastering The Interview	9240 non-null	object
36	Last Notable Activity	9240 non-null	object

Fig. 2. Features the dataset.

3.3 Pre-processing

There are some missing values in some features as you can see in Fig. 2. Some of them are part of the analysis but some are not like 'What is your current occupation', 'What matters most to you in choosing a course', Asymmetric Activity Index, Asymmetric Profile Index, Asymmetric Activity Score, and Asymmetric Profile Score so we dropped them. Further, we did some processing in the column Lead Source. For that, we transferred some of the values of the lead source into others, for example, there were two values Google and google so we converted them into a single value. Also, we transformed some values like YouTube and Facebook into social media. Then we filled in the missing values for this column with Google. In the next step, we replaced the null values of 'Page views per visit with the median. For the column 'Last Activity' we replaced null values with "Email opened" as there were fewer null values and value counts for "Email opened" were highest. In the country column, we replaced null values with India. For the specialization and 'How did you hear about X education' feature, we replaced null values and, 'select' values with Unknown. In the City column, we replaced values with Mumbai for the data that has India as a country and the rest as 'Other international city'. We replaced the null values of Total visits with a median.

After that, we made 3 columns from the features, Total visits, Total time spent on the website, and page views per visit which we derived by getting the square root of each value of those columns. Then we deleted the original columns as they were not required.

Further One Hot Encoder was applied to all the categorical features of the dataset which helped with the model building in this work.

3.4 Exploratory Data Analysis

To understand data better, we did some analysis of various features (Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16)

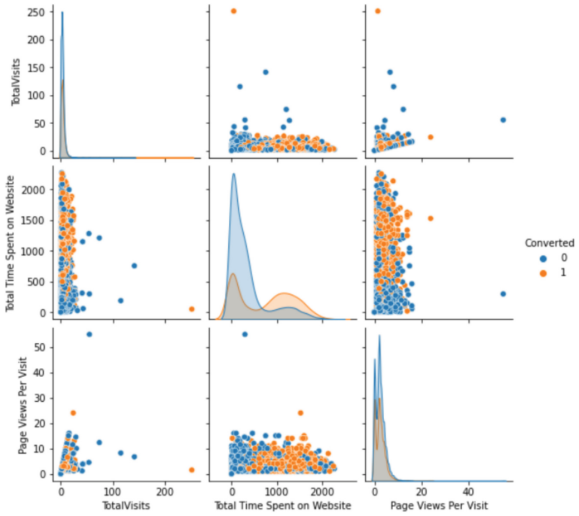


Fig. 3. Data Distribution of Total visits, Total time spent on the website, and Page views per visit.

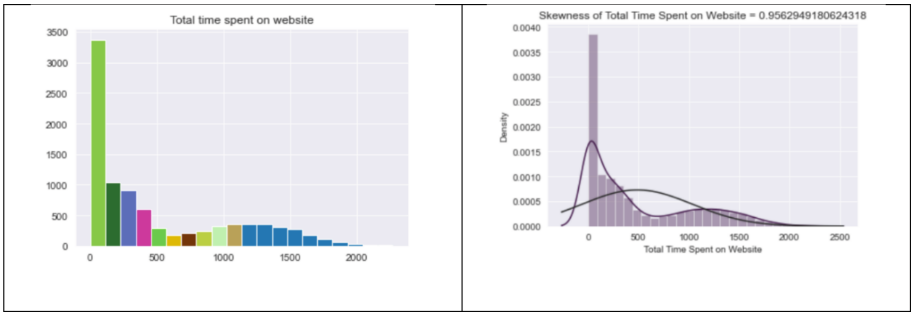


Fig. 4. Histogram and Skewness of Total Time Spent on the website by the customer respectively.

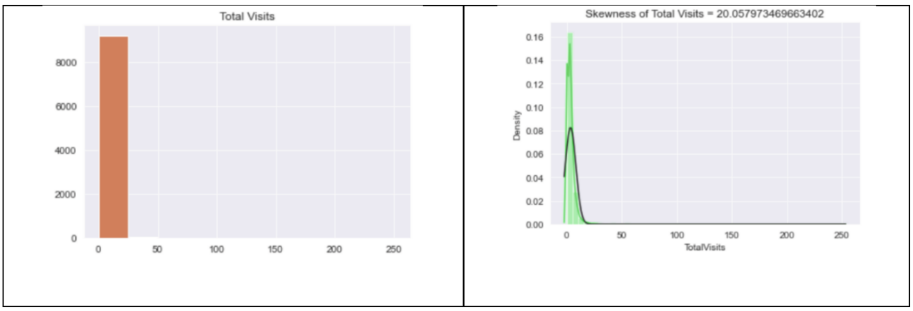


Fig. 5. Histogram and Skewness of Total Visits respectively.

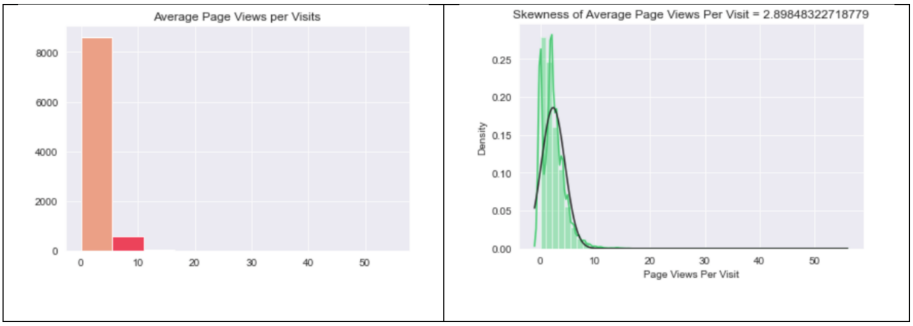


Fig. 6. Histogram and Skewness of Average Page Views per Visit respectively.

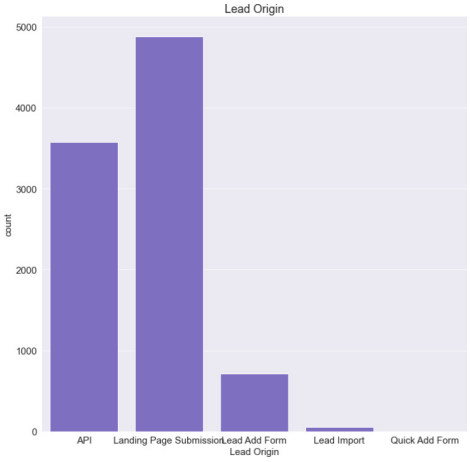


Fig. 7. Count of leads for each Origin.

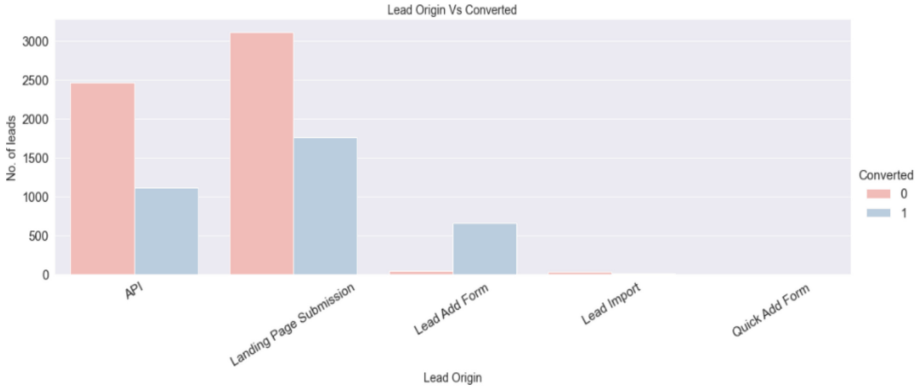


Fig. 8. Count of leads for each Origin and Converted.

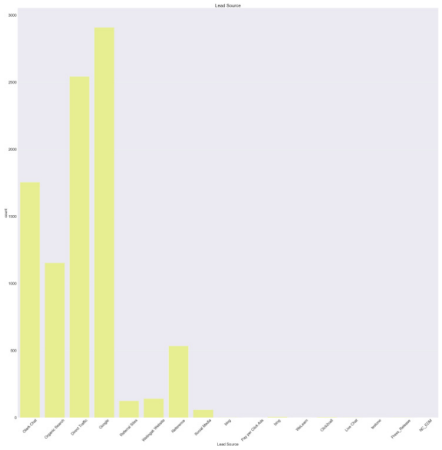


Fig. 9. Count of leads for each Source.

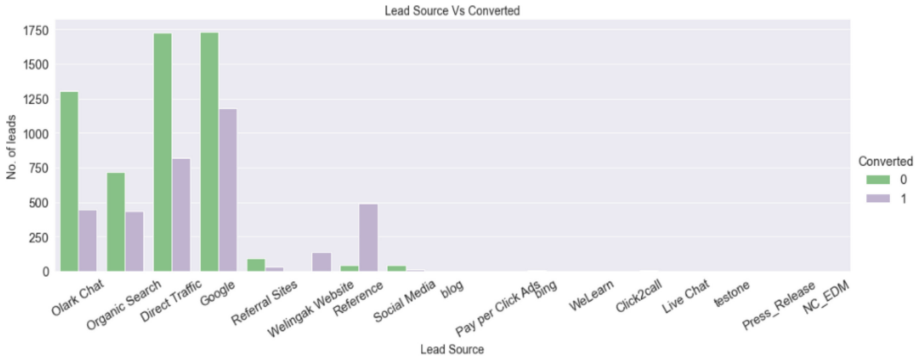


Fig. 10. Count of leads for each source and converted.

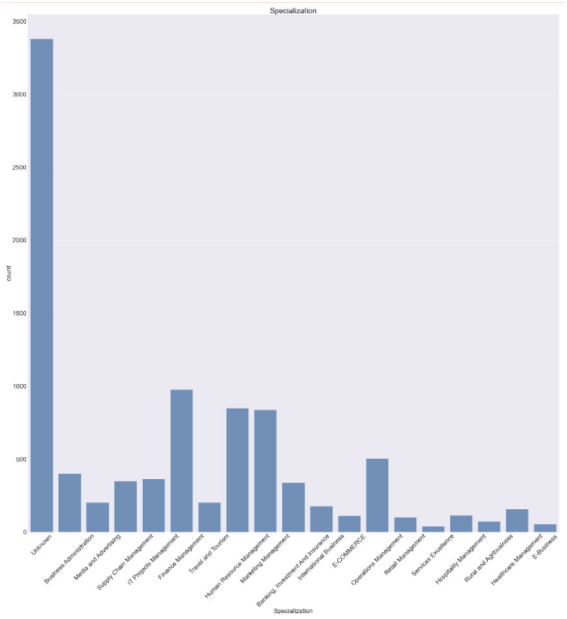


Fig. 11. Count of leads for each Specialization.

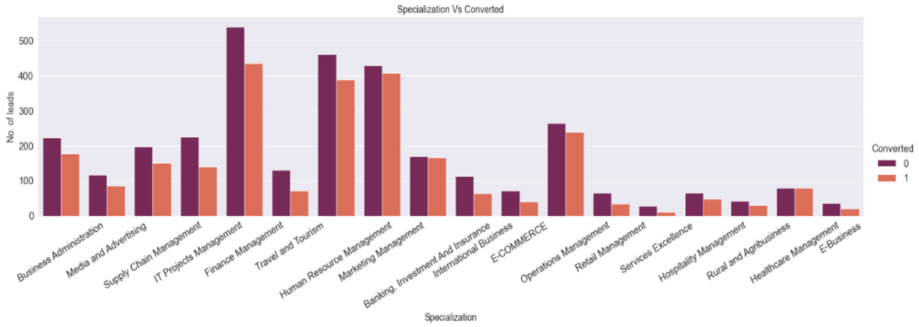


Fig. 12. Count of leads for each specialization and converted.

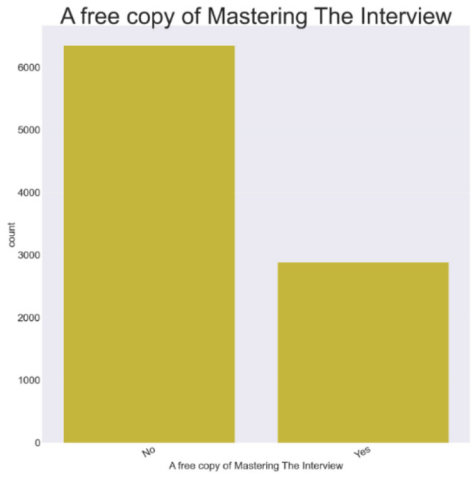


Fig. 13. Count of leads for each of A free copy of mastering the Interview.

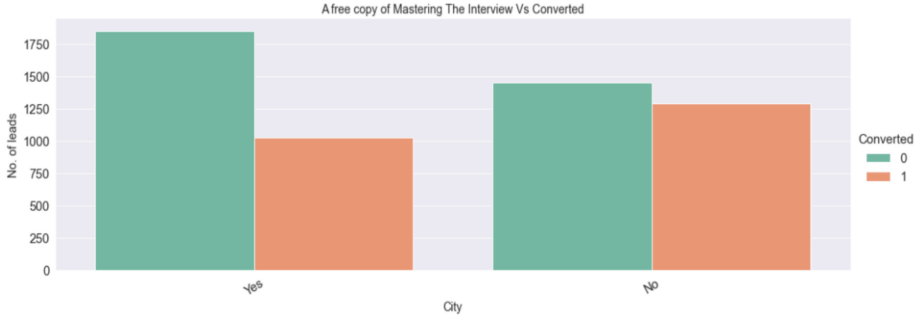


Fig. 14. Count of leads for each of A free copy of mastering the Interview and Converted.

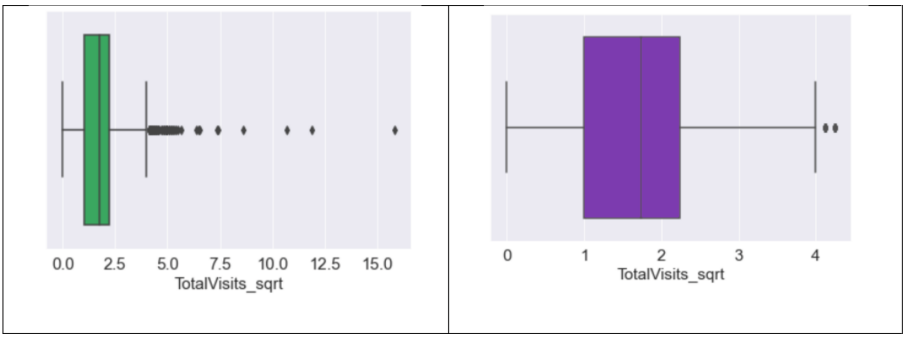


Fig. 15. Before vs After removing Outlier from Total Visits sqrt.

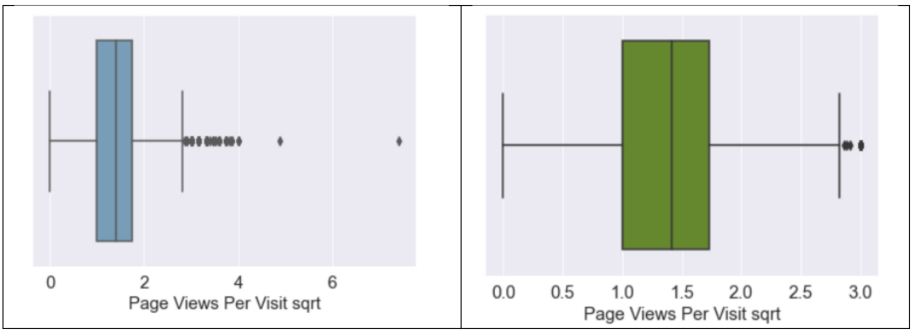


Fig. 16. Before vs After removing Outlier from Page views per Visit sqrt.

4 Design and Implementation

4.1 Design

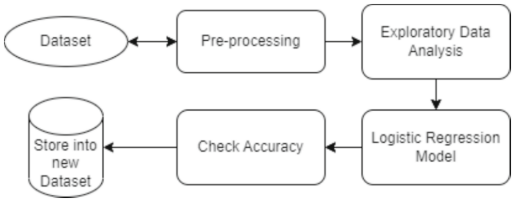


Fig. 17. Stages of Lead Conversion

As seen in Fig. 17 we applied some pre-processing and EDA on the data of leads of the company (Sect. 3). Further, we will be building a model to convert and nurture some potential leads to enhance the conversion ratio.

4.2 Logistic Regression

Contrary to the name, LR is a straightforward but incredibly powerful binary and linear categorizing technique. It is frequently utilized in experiments as well as in business wherever linear grouping is required. The approach uses a sigmoid function to transfer a latent characteristics vector to the range [0, 1]. An LR model can train more quickly than more complicated algorithms like Artificial Neural Networks (ANN) and Random Forests while still achieving respectable and comparable results. Researchers in many different scientific fields continue to find LR to be particularly intriguing because of its effectiveness and simplicity.

For this work, we used a logistic regression model to evaluate the conversion rate and modify the lead score so that the conversion rate can be higher. The dataset was split into 2 parts: training and testing. The target feature for this model is the Converted column. The importance of each column was extracted from the training set. Next, the logistic regression model was applied to the training dataset.

4.3 Model Evaluation

The lead conversion rate before building the model was very poor. To increase the rate model tuning was needed. Further, the training and testing set was modified using the feature of importance. Next, the Generalized linear model (GLM) was applied to the new training set. GLM is used for data that has a non-normal distribution, and it is a highly effective model for modeling variables with dependencies that have a non-normal distribution. After that VIF was calculated for each of the independent variables to check the correlation between them. While numerous independent variables are multicollinear, VIF quantifies the amount that the variability of a regression coefficient has increased owing to multicollinearity. Many features had low VIF scores. So, we chose a feature with a VIF of less than 5.

After updating the training and testing set for independent variables, again we applied the logistic regression model to the training dataset. It was observed that the accuracy was 75.32 which was increased from before. Next, we applied the GLM model again to the training set. Then, predicted values for the converted feature were stored in the final training set. The Roc curve was built in the next step based on an arbitrary predicted value of 0.5. The optimal probability cutoff point was derived, and the predicted column was modified in the final training dataset of the converted feature. Further, we stored the rounded converted probability score in the resulting data.

5 Result

Logistic regression was applied to the lead data of an X education company as seen in the Model evaluation of Sect. 4. In this section, we show the result of the model that we built. We will be discussing the result of the confusion matrix, roc-curve, and final accuracy.

```
Sensitivity(True positive rate):  0.6202414113277623
False Positive Rate:  0.16224188790560473
Specificity(True negative rate):  0.8377581120943953
False Negative Rate:  0.3797585886722377
Precision:  0.7083775185577943
F-Score:  0.6613861386138614
```

Fig. 18. Confusion matrix

5.1 Confusion Matrix

As seen in Fig. 18 we got the results of a True Positive rate of 0.62, False Positive rate of 0.16, True negative rate of 0.84, False negative rate of 0.38, precision = 0.70, and F-score is 0.66. Next, we built the ROC curve.

5.2 ROC-Curve

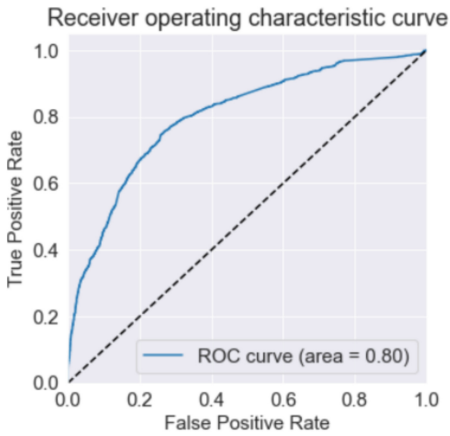


Fig. 19. ROC-curve

As seen in the above figure ROC curve was built on the value of False Positive rate, True positive rate, and accuracy score. From the graph, it can be observed that the accuracy score is 0.80 (Fig. 19).

5.3 Final Result

After getting the ROC-curve optimal cut-off point was derived from plotting the line graph.

We built the line graph from the measures of accuracy, sensitivity (True Positive rate), and specificity (True negative rate). From the graph, it can be observed that the optimal point is approx. 0.37. Using the optimal Predicted value, we derived the value predicted in the training data of Converted (Fig. 20).

Further, the conversion rate was checked from the Resulting data of converted leads. The conversion rate of the lead score was found to be 70%. And we stored the final lead scoring dataset in the CSV file.

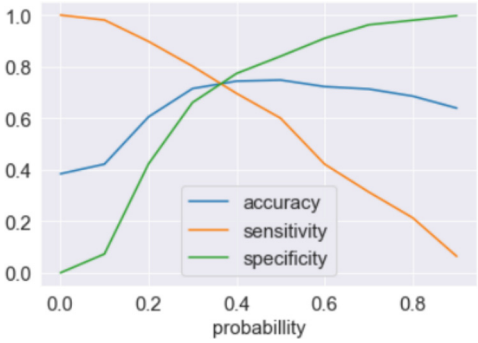


Fig. 20. Line graph for optimal probability point

6 Conclusion

In this work, the recommended lead qualifying framework was provided for developing a lead rating system in a reduced form for small businesses. It determines various parameters for calculating lead scores and uses a simple spreadsheet to reduce the model's complexity. The characteristics, such as lead origin, total time spent on the website, total visits, specialization, and last notable activity are critical and have a significant impact on lead scores. A solid lead score model will give a path for the firm to boost sales productivity and effectiveness. As a result, by emphasizing important data and giving importance to the conversion rate the lead score can be increased. In this work, a regression model was built to increase the conversion rate. The final conversion rate was 70% which we observed increased from the last conversion rate of 30%. Further, we can enhance this rate by applying models like using Naïve Bayes and Random Forest and comparing all results from these 3 models.

References

Banerjee, S., Bhardwaj, P.: Aligning marketing and sales in multi-channel marketing: compensation design for online lead generation and offline sales conversion. J. Bus. Res. **105**(June), 293–305 (2019). https://doi.org/10.1016/j.jbusres.2019.06.016

Benhaddou, Y., Leray, P.: Customer relationship management and small data - Application of Bayesian network elicitation techniques for building a lead scoring model. In: Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 2017-October, 251–255 (2018). https://doi.org/10.1109/AICCSA.2017.51

Diogo, J., Silva, D., Rodrigues, S.: Automated lead scoring system: a case study of a Portuguese startup (2020)

Duncan, B., Elkan, C.: Probabilistic modeling of a sales funnel to prioritize leads. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015-August, 1751–1758 (2015). https://doi.org/10.1145/2783258.2788578

Etminan, A.: Prediction of Lead Conversion With Imbalanced Data: A Method based on Predictive Lead Scoring (2021). www.liu.se

Indexed, S.: Implementing lead qualification **11**(10), 81–90 (2020). https://doi.org/10.34218/IJM.11.10.2020.008

Jadli, A., Hamim, M., Hain, M., Hasbaoui, A.: Toward a smart lead scoring system using machine learning. Indian J. Comput. Sci. Eng. **13**(2), 433–443 (2022). https://doi.org/10.21817/indjcse/2022/v13i2/221302098

Kumar, G.N., Hariharanath, K.: Designing a lead score model for digital marketing firms in education vertical in India. Indian J. Sci. Technol. **14**(16), 1302–1309 (2021). https://doi.org/10.17485/ijst/v14i16.290

Lindahl, E., Skolan, K., Datavetenskap, F., Kommunikation, O.: A qualitative examination of lead scoring in B2B marketing automation, with a recommendation for its practice (2017). http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-213432

Monat, J.P.: Industrial sales lead conversion modeling. Mark. Intell. Plan. **29**(2), 178–194 (2011). https://doi.org/10.1108/02634501111117610

Nygård, R., Mezei, J.: Automating lead scoring with machine learning: an experimental study (2020). In: Proceedings of the Annual Hawaii International Conference on System Sciences, 2020-January, 1439–1448. https://doi.org/10.24251/hicss.2020.177

Wu, M., Andreev, P., Benyoucef, M.: The state of lead scoring models and their impact on sales performance. Inf. Technol. Manage. (2023). https://doi.org/10.1007/s10799-023-00388-w

Artificial Intelligence in Gestational Diabetes Mellitus

Amna Kausar(✉), Shravani Kulkarni, Piyush Bhosale, Susanta Das, and Khushbu Trivedi

Ajeenkya DY Patil University, Pune, India
2004amnas@gmail.com

Abstract. The following article reviews existing studies on Gestational Diabetes Mellitus (GDM) and the advancements made in Artificial Intelligence (AI) to support healthcare professionals and patients. Our article briefs over the health risks associated with GDM for the mother and fetus as well as the diagnosis processes involved. In our paper, we also discuss the various AI methods such as GBDT, ANNs and Deep Learning (DL) algorithms like LSTM (Long Short-Term Memory) used in recent studies with their corresponding performance scores. We also mention the methodology we have followed in obtaining our sources for both data and algorithms from open-source platforms such as Kaggle. Following this methodology, we have curated various sources for GDM datasets that have been used in studies and can be used for further development in AI. Throughout the article, we consistently focus on the lack of information and development of modern technology in women-focused areas. In addition, our article aims to bring awareness of using modern technology for prediction of female diseases, such as GDM.

Keywords: GDM · AI · Deep Learning · datasets · algorithms · performance

1 Introduction

The paper aims to address the usage of Artificial Intelligence (AI) in Gestational Diabetes Mellitus (GDM), an often-overlooked area of modern technology used to classify women-specific diseases. Copious research has been conducted for identifying and predicting the presence of type 1 and 2 diabetes. However, even though there are certain research conducted consisting of only women participants, it is of diseases present in both genders. For instance, the PIMA diabetes dataset, one of the most widely used datasets worldwide for data science and ML practices although features women, does not take attributes specific to women or address the presence of GDM- A leading factor in what can increase the likelihood of developing permanent diabetes.

Glucose and insulin levels are two metrics that are closely monitored during pregnancy. A healthy pregnant woman has glucose levels of 140 mg per deciliter (mg/dL) and lower. Gestational Diabetes Mellitus (GDM) occurs when there is an imbalance in glucose levels which results in high blood sugar levels, affecting up to 20% of all pregnancies worldwide [1].

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 131–137, 2026.
https://doi.org/10.1007/978-3-031-98138-8_10

The effect is not permanent in all cases however and could lead to type 2 diabetes if left untreated later down the road. The effects of GDM can be detrimental to both the mother and baby, putting importance on the need for early detection and monitoring for the wellbeing of both individuals [2].

We understand that due to the lack of efforts put into women-focused research since the early days, not much progress has been made for GDM specifically. We also aim to introduce interpretable models, addressing the issue of the black box in deep learning to better understand and aid researchers in investigating how certain attributes may influence the outcome.

In addition, our article addresses the lack of diversity in ethnicities. Different regions have different thresholds for what is deemed average or abnormal. By training the model on diverse data with their ethnicities mentioned, we can better understand and iterate on the classifier to improve the accuracy.

Machine Learning (ML) and Artificial Intelligence (AI) have been advancing rapidly in the field of healthcare, and our article aims to spread awareness and enlighten readers on recent developments occurring within AI and how they can be applied to addressing GDM by synthesizing existing literature and pointing to new directions for the future of GDM management.

2 Understanding GDM

GDM is commonly screened for between 24 to 28 weeks of pregnancy who have no pre-existing cases of diabetes, usually diagnosed by using an Oral Glucose Tolerance Test (OGTT) where the fasting plasma glucose levels are measured for 1, 2 and 3 h. The diagnosis is complete if two of the following parameters are met and exceeded [3]. Table 1 displays the different thresholds used in determining the completion of a diagnoses.

Table 1. Plasma Glucose Levels for GDM diagnoses

Type	Plasma Glucose Levels (mg/dL)
Fasting	≥92
1-h	≥180
2-h	≥153

2.1 Risks

Factors such as age, obesity, number of pregnancies, family history and even ethnicity play important roles in the development of this condition and can push forward the importance of being under supervision much earlier as their risk of having undiagnosed type 2 diabetes is higher [4].

Risks associated with GDM are shown in the Tables 2 and 3 for both the mother and the fetus. Appropriate prenatal care must be done in order to protect the wellbeing of the mother and her unborn child [5, 6].

Table 2. Maternal health risks

Risk	Description
Preeclampsia	Increased blood pressure leads to complications for both
Cesarean Delivery	Higher chance of requiring a C-section due to fetal macrosomia (large baby)
Postpartum Diabetes	Women with GDM have a 7-fold increased risk of developing type 2 diabetes later in life
Gestational Hypertension	Increased blood pressure leads to complications for both
Infection	Higher susceptibility to infections during and after pregnancy
Uterine Atony	Risk of postpartum hemorrhage after delivery
Prolonged Labor	Increased likelihood of prolonged labor due to complications associated

Table 3. Fetal health risks

Risk	Description
Macrosomia	Excessive fetal growth can lead to complications during delivery, such as shoulder dystocia
Neonatal Hypoglycemia	Babies born to mothers with GDM may experience low blood sugar levels after birth
Preterm Birth	Increased risk of early delivery leads to health complications for the infant
Respiratory Distress Syndrome	Newborns may experience breathing difficulties due to immature lungs
Stillbirth	Higher risk of stillbirth

3 AI Algorithms

3.1 Gradient Boosting Decision Trees (GBDT)

GBDT consists of a model that builds sequentially on weak learners by correcting previous mistakes and resulting in a stronger model. In one study, it outperformed traditional models such as logistic regression and Supervised Vector Machine (SVM) with an Area Under Curve (AUC) score of 0.67 for certain groups. The study was based on Japanese participants [7].

3.2 Artificial Neural Network (ANN)

ANN is a model inspired from biological processes occurring in a brain cell. It can capture more complex patterns present in data. The MIDO model uses ANN as the primary approach for GDM prediction using best predictive variables for GDM risk including age, BMI, family history and so on. The model reached an AUC score of 0.8471. The study was based on Mexican participants [8].

3.3 Deep Learning

Deep Learning (DL) is a subset of AI and ML that makes use of ANNs to mimic the processes in humans. It derives patterns from unlabeled data to make predictions. Techniques such as LSTM (Long Short-Term Memory) and RNN (Recurrent Neural Networks) have been used in a study to produce a model that could cut down on cost and time by reducing the need for OGTT for patients that do not have a high risk of GDM. The model gave an AUC score of 0.98 [9]. LSTM is typically used for time series data and is a type of RNN. They're designed in such a way that they retain information from previous steps in their architecture. This makes LSTMs extremely useful in other areas such as NLP (Natural Language Processing) for contexts. RNNs in general are a type of ANN and consist of feedback loops which makes it better at recognizing hidden patterns, a characteristic consistently observed in DL algorithms.

4 Data Utilization

4.1 Quality

Predictive models rely on data that is complete, accurate and does not contain missing values. They need to represent the entirety of population as even the slightest change can lead to inaccurate predictions. Data should be cleaned to impute missing values and maintain a standard protocol during data collection.

4.2 Breadth

A larger set of features is required to capture the intricacies involved when dealing with GDM. Since there are multiple variables such as ethnicity, age, BMI etc., data that is representative of lifestyle choices should be carefully considered during data collection [10].

4.3 Sources

Data should be sourced from clinical organizations that provide real health data that is publicly available or sourced privately with consenting patients. National health records provide extensive datasets that may give insights to patterns associated with diverse populations. Electronic Health Records (EHRs) may provide timeseries datasets that allow real-time tracking of indicators. These records may also include patient history on existing conditions [11]. Table 4 directs to papers that have used various different sources for training which can be accessed by reviewing the papers listed.

Table 4. Data sources with brief descriptions and papers that have used it.

Source	Description	Papers
National Vital Statistics System (NVSS)	Provides birth certificate data and GDM likelihood estimates	[12]
Gene Expression Omnibus (GEO)	Contains gene expression data related to GDM	[13]
BMC Medicine	Population based cohort study that developed and validated prediction models for GDM treatment modalities using ML	[14]
Malaysian SECOST cohort	Study investigating GDM in Malaysian pregnant women	[15]
Growing Up in New Zealand Study	Contain data on GDM effects on child development	[16]

5 Methodology

A systematic review was conducted using Google Scholar as our primary search engine. Our search queries consisted of "AI in GDM prediction", "diabetes classifiers", "ensemble learning" and other combinations of the keywords Artificial Intelligence, Machine Learning, Classifiers, GDM, and diabetes management system. The results consisted of various articles, journals, and research papers both technical and healthcare-focused.

We also obtained datasets from literature implementing their datasets as displayed in Table 4 which in addition to providing datasets also gave an idea of the algorithms being used for classifying and predicting GDM.

In addition to reviewing previous literature, we leveraged the publicly available datasets from Kaggle to gain a better understanding of the lack of relevant data consisting of primary attributes associated with women such as the number of pregnancies, hormonal imbalances, and more. The keyword GDM dataset was used to filter out relevant datasets revealing unsurprisingly low-quality data. These crucial attributes are often overlooked much less recorded and available for usage.

6 Conclusion

The integration of AI in management of GDM has advanced the lives of many, including healthcare professional and patients alike. As a result, there have been notable improvements on prenatal care and timely detection at the time of pregnancy. AI techniques have evolved from traditional models to more complex methods such as deep learning, using ensemble methods and applying optimization algorithms to improve model performance. However, further research is essential to cater to diverse populations and develop models that can reach their full potential to improve the experiences of women throughout their pregnancies. To do this, individuals may need to access highly confidential health records that are not readily available to the public. In addition, the data

provided by hospitals may be difficult to standardize and may be a daunting task to clean and use for developing models, a framework must be put into place. Our article brings together some well-known AI algorithms, presenting the lack of insights into female diseases and giving a real-life problem as an opportunity to those who can contribute to the cause. Thus, our article gives a head start to seasoned developers and researchers in this field of expertise. We encourage individuals and organizations to bridge the gap by utilizing developing technologies including but not limited to DL algorithms, meta classifiers and quality datasets for the betterment of what makes up the other half of the world.

References

1. Alfadhli, E.M.: Gestational diabetes mellitus. Saudi Med. J. **36**(4), 399–406 (2015). https://doi.org/10.15537/smj.2015.4.10307
2. Alduayji, M.M., Selim, M.: Risk Factors of Gestational Diabetes Mellitus Among Women Attending an Antenatal Care Clinic in Prince Sultan Military Medical City (PSMMC), Riyadh, Kingdom of Saudi Arabia: A Case-Control Study. Cureus. https://doi.org/10.7759/cureus.44200
3. Karagiannis, T., Bekiari, E., Manolopoulos, K., Paletas, K., Tsapas, A.: Gestational diabetes mellitus: why screen and how to diagnose (2010). https://pmc.ncbi.nlm.nih.gov/articles/PMC2943351/
4. Virjee, S., Robinson, S., Johnston, D.G.: Screening for diabetes in pregnancy. J. R. Soc. Med. **94**(10), 502–509 (2001). https://doi.org/10.1177/014107680109401003
5. Muche, A.A., Olayemi, O.O., Gete, Y.K.: Effects of gestational diabetes mellitus on risk of adverse maternal outcomes: a prospective cohort study in Northwest Ethiopia. BMC Pregnan. Childbirth **20**(1) (2020). https://doi.org/10.1186/s12884-020-2759-8
6. Nakshine, V.S., Jogdand, S.D.: A comprehensive review of gestational diabetes mellitus: impacts on maternal health, fetal development, childhood outcomes, and long-term treatment strategies. Cureus (2023). https://doi.org/10.7759/cureus.47500
7. Watanabe, M., et al.: Prediction of gestational diabetes mellitus using machine learning from birth cohort data of the Japan Environment and Children's Study. Scientific Repor. **13**(1) (2023). https://doi.org/10.1038/s41598-023-44313-1
8. Gallardo-Rincón, H., et al.: MIDO GDM: an innovative artificial intelligence-based prediction model for the development of gestational diabetes in Mexican women. Sci. Reports **13**(1) (2023). https://doi.org/10.1038/s41598-023-34126-7
9. Kurt, B., et al.: Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. Med. Biol. Eng. Compu. **61**(7), 1649–1660 (2023). https://doi.org/10.1007/s11517-023-02800-7
10. Zhang, Z., et al.: Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J. Med. Internet Res. **24**(3), e26634 (2021). https://doi.org/10.2196/26634
11. Cubillos, G., et al.: Development of machine learning models to predict gestational diabetes risk in the first half of pregnancy. BMC Pregn. Childbirth **23**(1) (2023). https://doi.org/10.1186/s12884-023-05766-4
12. Bolduc, M.L., et al.: Gestational diabetes prevalence estimates from three data sources, 2018. Matern. Child Health J. **28**(8), 1308–1314 (2024). https://doi.org/10.1007/s10995-024-03935-1
13. Binder, A.M., LaRocca, J., Lesseur, C., Marsit, C.J., Michels, K.B.: Epigenome-wide and transcriptome-wide analyses reveal gestational diabetes is associated with alterations in the human leukocyte antigen complex. Clinical Epigen. **7**(1) (2015). https://doi.org/10.1186/s13148-015-0116-y

14. Liao, L.D., et al.: Development and validation of prediction models for gestational diabetes treatment modality using supervised machine learning: a population-based cohort study. BMC Med. **20**(1) (2022). https://doi.org/10.1186/s12916-022-02499-7
15. Yong, H.Y., Shariff, Z.M., Rejali, Z., Yusof, B.N.M., Yasmin, F., Palaniveloo, L.: Seremban Cohort Study (SECOST): a prospective study of determinants and pregnancy outcomes of maternal glycaemia in Malaysia. BMJ Open **8**(1), e018321 (2018). https://doi.org/10.1136/bmjopen-2017-018321
16. Lawrence, R.L., Wall, C.R., Bloomfield, F.H.: Prevalence of gestational diabetes according to commonly used data sources: an observational study. BMC Pregn. Childbirth **19**(1) (2019). https://doi.org/10.1186/s12884-019-2521-2

A Comprehensive Risk Assessment Framework for Multiple Natural Hazards Using CLIMADA Model

Jaya Nidhi Kandir, Gauri Deshpande(✉), and T. P. Singh

Symbiosis Institute of Geoinformatics, Symbiosis International (Deemed) University, Pune, India
gaurideshpande127@gmail.com

Abstract. In modern eras, the effect of natural hazards has become more evident due to climate change, which is inconsistent in terms of its intensity, frequency and geographical distribution of the occurrences. Thus, it was thought necessary to evaluate and forecast the impacts of natural hazards under various climate scenarios through sophisticated tools and models. One of tools is the CLIMADA (CLIMate ADAptation) model, which is available in open domain, deals with probabilistic risk assessment intended to compute the risks and impacts of natural hazards, in view of socio-economic and climatic factors. The principal goal of this research is to employ the CLIMADA model to conduct a comprehensive risk assessment of extreme heat related natural hazards under various climate change scenarios. As per the projections done by Representative Concentration Pathways (RCP) scenario, the investigation is performed on the basis of changing climate variables and its impact on extreme heat conditions. It not only provides the details of intensity but also explains how much time it will occur again. Ultimately, the research goals to deliver scientific understandings into future risks that facilitate communities, governments, and organizations to develop sustainable and adaptive measures to survive with the impressions of climate change.

Keywords: CLIMADA · climate change · natural hazards

1 Introduction

Natural hazards are prevailing and often shattering events that have potential to affect the lives, property and other assets which causes disruption to ecosystems and economies. Though these events are natural, its severity depends on the involvement of the human population. In recent years, research findings published by the Intergovernmental Panel on Climate Change (IPCC), has focused on extreme weather events that are increased due to climate change. This is an alarming because intensity, frequency and geographical expansion of these events were altered. These extreme weather conditions (heatwaves, cold waves, heavy rainfall etc.) are become hazards that harmful for the life on the earth's surface.

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 138–148, 2026.
https://doi.org/10.1007/978-3-031-98138-8_11

Major collaborator of green gases in the troposphere is anthropogenic activities causing climate change. These are foremost activities which adversely affects the weather pattern on the global scale. The changing weather patterns are not consistent and erratic in nature for example, global temperatures are rising that affects water cycle which attributed to excess rainfall and severe droughts. These severe draughts also create situation which benefits for the wildfire. Again with such drastic changes in the weather pattern leads to severe development of hurricanes and typhoons. Considering the variability in extreme climatic events in the decade, it is crucial to understand its direct and indirect impacts on climate and forecast the changes. Intergovernmental Panel on Climate Change (IPCC) has established Representative Concentration Pathways (RCPs). This initiative is to study the green house effects on different climate conditions. Basically, these climate projections depend on a range of factors and assumptions that replicate systematic knowledge of the climate system, greenhouse gas (GHG) emissions, and socio-economic trends. These elements form the basis for estimating future changes in temperature, precipitation patterns, sea levels, and the frequency and intensity of extreme weather events. With help of the studied/ accumulated data eventually one can provide significant predictions on weather conditions. For most of the projections, pre-industrial climate (1850–1900) is considered as the baseline. Future GHG emissions are clustered into predefined scenarios, for example, RCP2.5 (minimum GHG emissions) to RCP8.5 (maximum GHG emissions). SSPs combine socioeconomic conditions with RCPs to reflect plausible futures. Bearing in mind these assumptions, climate projections are typically prepared for short-term (2020–2040), mid-term (2040–2060), and long-term (2080–2100) periods.

As mankind makes progress on study and analysis of this data one will be able to understand and provide possible solutions to the future disaster which can majorly benefit disaster management departments of government officials. It will create awareness on possible variety of skills and tools that provide baseline guidance for sustainable human life. With the generation of AI tools/technology guidance, one can create custom data models that can predict and provide valuable feedbacks that benefit lives on the earth. CLIMADA is one of the models that is used to compute the risks and impacts of natural hazards with the consideration of climatic as well as socio-economic factors. By integrating climate projections from RCP scenarios with data on exposure and vulnerability, CLIMADA can estimate potential damage and losses from natural hazards. This model supports policymakers, planners, and stakeholders in making informed decisions about adaptation and mitigation strategies. The primary goal of present research is to utilize the CLIMADA model to conduct a comprehensive risk assessment of hazards related to extreme heat under different climate change scenarios.

2 Review of Literature

Since 1990, natural disasters have claimed the lives of approximately 1.6 million people globally and caused an estimated USD 260 billion to USD 310 billion in economic losses [1, 2]. Over the past fourteen years, the scientific and policy communities have hastened the development of global models to estimate the risks associated with natural hazards on a global scale, realizing the urgency of mitigating these risks. The scientific

literature on global-scale natural hazard risk assessments is thoroughly re- viewed in this work, with particular attention on how future forecasts of hazards, exposure, and vulnerability are incorporated into the assessments. The international organization in charge of evaluating the science behind climate change is the Inter- governmental Panel on Climate Change (IPCC). In 1988, the IPCC provided policy- makers with information on: the scientific study of climate change's effects, potential dangers in the future, and strategies for mitigation and adaptation.

To investigate future climate change and its effects, two complimentary frameworks are used: the Shared Socioeconomic Pathway (SSP) and the Representative Concentration Pathway (RCP). Relative to pre-industrial levels, varying amounts of radiative forcing (measured in watts per square meter) by the year 2100 are described by greenhouse gas concentration trajectories, or RCPs. The four primary RCP scenarios, which range from low to high future climate change, are RCP2.6, RCP4.5, RCP6.0, and RCP8.5. CMIP6 global climate and meteorological datasets were the primary data source for this project. Kumar et al. (2023) conducted a study to analyze bias-corrected simulations of bias-corrected General Circulation Models (GCMs) from the CMIP6 in India's future climate. For the near future (2021–2050) and the far future (2071–2100), these models were used to project changes in mean rainfall and daily extremes, such as the number of rainy days and daily intensity, during the southwest (SW) and northeast (NE) monsoon seasons, relative to the baseline period (1985–2014), under the SSP2-4.5 and SSP5-8.5 scenarios. The study also looked at variations in daily extremes of mean, minimum, and maximum temperatures [3]. Dawkins L.C. et al. (2023) have applied both CLIMADA and ensembles Generalised Additive Model for assessing the climate risk for the United Kingdom (UK). As they have adopted combined approach that is open source CLIMADA and ensemble model to identify the risks, which is inexpensive and reliable [4]. In the changing climate scenario, the assessment of global multiple hazard risk is feasible using open source framework of CLIMADA through various case studies related to tropical cyclone and river floods by Stalhandske Z et al. (2024) [5]. Thus, it was thought necessary to understand the climate risk through CLIMADA for various hazards. This research illustrates the risk assessment related to extreme heat.

3 Objectives

The major objectives of the research are as follows:

3.1. To analyze the concept of Shared Socioeconomic Pathways (SSPs) and Representative Concentration Pathways (RCPs) in assessing the impacts of climate change on natural hazards.
3.2. To find out the best suited method to downscale CMIP6 climate raster datasets through literature review and then perform the downscaling.
3.3. To analyze historical data and climate projections to identify trends and future changes in the frequency, intensity, and spatial distribution of specific natural hazards.
3.4. Utilize CLIMADA and other tools to assess the potential impacts of increased natural hazard risk on the study region, considering economic losses, infrastructure damage, and ecological disruption.

4 Study Area

The western coast of India comprises the states of Gujarat and Maharashtra, which have been selected as the study area for this research because of their unique geographical characteristics and susceptibility to extreme heat conditions. Especially during the pre-monsoon months, Gujarat and Maharashtra both suffer from severe heatwaves. These states' interior regions frequently record temperatures that rise above 40 °C, endangering public health and placing a strain on water and electricity supplies.

5 Data and Methods

The methodology adopted for the study is depicted in Fig. 1.

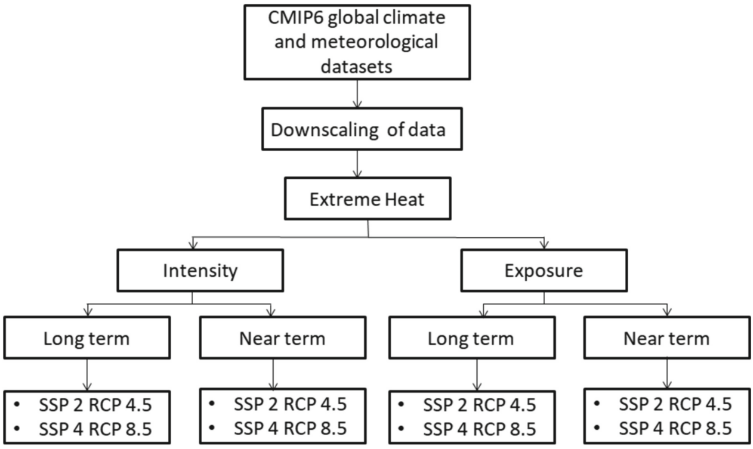


Fig. 1. Methodology

The inputs for the present model are categorized into four sections: i) Hazard data: To the hazard data, CMIP6 data for global climate and meteorological datasets were applied. The Coupled Model Intercomparison Project Phase 6 (CMIP6) represents the forefront of global climate model simulations, offering crucial insights into the Earth's climate past, present, and future [6]. ii) Exposure data: Spatial distribution of assets, population, or infrastructure at risk data is applied to show the variation in risk intensity and probable impact iii) Climate scenarios: this incorporates data aligned with Representative Concentration Pathways (RCPs) or Shared Socioeconomic Pathways (SSPs) and allows for assessing impacts under different climate futures.

Global climate models (GCMs) like those in CMIP6 typically operate at coarse spatial resolutions, often in the range of 100–250 km. This resolution is insufficient for capturing fine-scale climate processes and variability that are crucial for regional and local applications. Downscaling enhances spatial resolution, providing more detailed and precise climate information that can be better used for local decision-making and planning.

5.1 EBK Regression Prediction

The raster data from IPCC GWI Interactive Atlas was imported into ArcGIS Pro and the following methodology (Fig. 2) was used for downscaling the raster for better calculation of results.

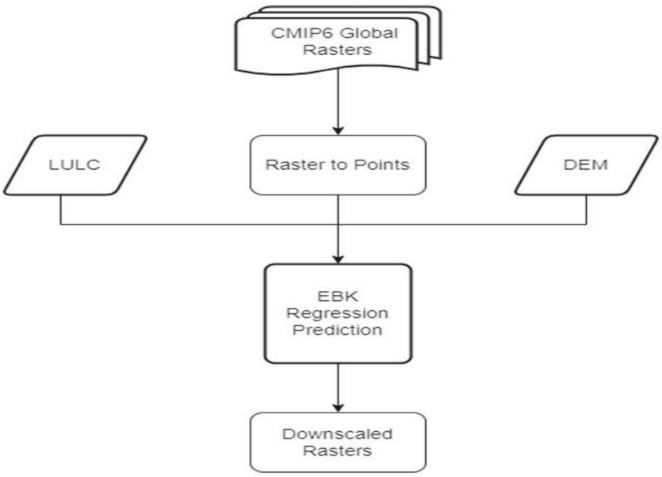


Fig. 2. CMIP6 downscaling Methodology

The EBK Regression Prediction (EBKRP) tool in ArcGIS Pro leverages its capability to combine regression analysis with spatial interpolation. High resolution LULC and DEM were used as explanatory or predictor variables. The LULC was from Sentinel-1 and 2 data with 10 m resolution. The DEM was from NASA SRTM v3 with 30 m resolution.

The Minimum Cumulative Percent of Variance, the parameter which specifies the minimum percentage of the total variance in the predictors was set to 95%. This parameter ensures that enough principal components are included to capture a significant portion of the variance in the predictor variables, which helps improve the accuracy and robustness of the regression model. Until the cumulative variance approaches or surpasses the designated Minimum Cumulative Percent of Variance, the tool adds up the variance described by the primary components one at a time. More principal components guarantee that a greater percentage of the variation is recorded, which could result in forecasts that are more accurate. It might, however, also result in an increase in computing load and model complexity.

As a result, comparatively higher resolution data with predicted values is obtained for further processing. The difference between the data before and after downscaling can be observed in the image given in Fig. 3.

5.2 Hazard Assessment in CLIMADA

CLIMADA is equipped to model a wide range of climate-related hazards, including tropical cyclones, river floods, droughts, and others. This flexibility allows users to assess

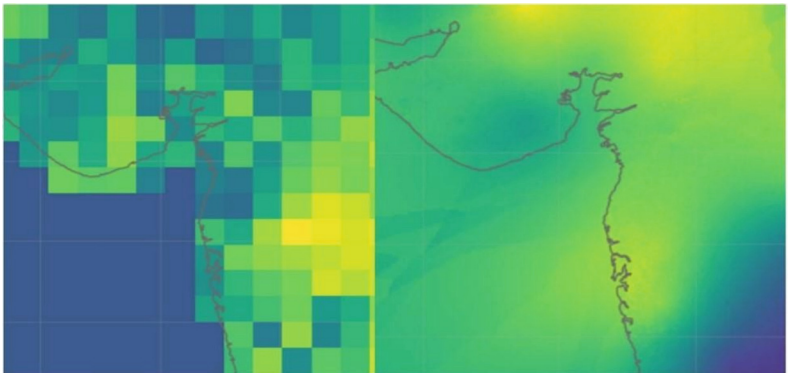


Fig. 3. Raw and downscaled data comparison

the impacts of different types of natural disasters on various assets and populations. For each hazard type, CLIMADA provides the capability to utilize predefined datasets, such as historical records of tropical cyclones, which can be easily integrated into the analysis. Additionally, users have the option to create custom hazard datasets tailored to specific regions or scenarios, enabling more precise and context-specific risk assessments [7]. This dual approach ensures that CLIMADA can be adapted to a broad spectrum of use cases, from global-scale studies to localized evaluations of climate risk. For current case study, a user-defined hazard function for extreme heat was taken. This function typically involves intensity, duration and frequency of extreme heat events. As an input, gridded temperature data, threshold value for extreme heat and events lasting 3+ days with historical probabilities derived from meteorological data.

5.3 Exposure Assessment in CLIMADA

A key element of climate risk assessments in CLIMADA are exposure objects, which offer comprehensive and configurable information on the assets that may be impacted by climate-related hazards. For precise impact modeling and wise decision-making in the planning of climate adaptation, this comprehensive data is essential. For this case study, exposure was taken as a distribution of assets all over the states of Gujarat and Maharashtra just to show the variation in risk intensity and probable impact (Fig. 4).

5.4 Vulnerability Assessment in CLIMADA

In CLIMADA, vulnerability functions are crucial instruments that measure the correlation between a hazard's intensity and the ensuing loss or damage to exposed assets. The way that various asset classes—such as buildings, infrastructure, and crops— respond to different degrees of danger intensity—such as wind speed, flood depth, and temperature extremes—is represented mathematically by these functions. Every vulnerability function captures the distinct qualities and susceptibilities of the assets concerned and is tailored to a specific danger type and asset category. For example, a vulnerability function taking into consideration construction standards and building materials would

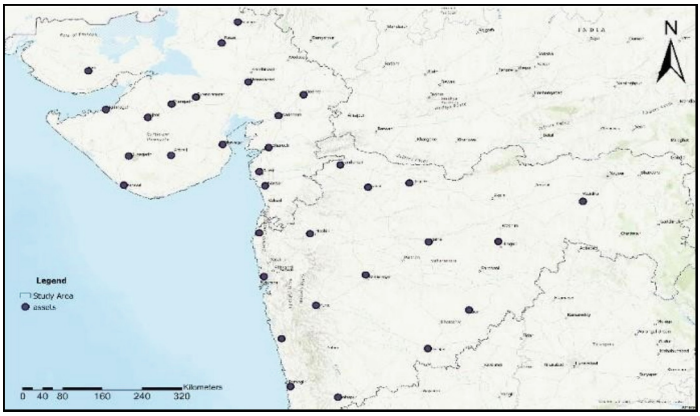


Fig. 4. Exposure assets for impact calculation

show projected damage levels at various wind speeds for residential buildings exposed to tropical storms. These capabilities allow CLIMADA to assess possible losses and damage under various hazard scenarios.

5.5 Impact Assessment in CLIMADA

An impact function is a key element of CLIMADA that quantifies the possible effects of climate related hazards on different assets by combining exposure data, vulnerability information, and hazard intensity. In essence, an impact function incorporates the vulnerability of the exposed assets and models the relationship between the intensity of a hazard and the consequent loss or damage to those assets. It estimates how different levels of hazard intensity (e.g., wind speed, flood depth) translate into damage (e.g., economic loss, physical destruction) using predefined or user-defined functions.

6 Results and Discussion

6.1 Extreme Heat

For the risk assessment of extreme heat using CLIMADA, global mean temperature and maximum temperature rasters for historical periods and future scenarios under RCP 4.5 and 8.5 were utilized. A hazard object was created in CLIMADA for extreme heat hazard type representing the intensity and frequency of extreme heat events. Subsequently, a custom impact function was defined based on the India Meteorological Department (IMD) definitions of a heat wave and severe heat wave. According to the IMD, a heat wave is characterized by a departure from normal temperature of 4.5 °C to 6.4 °C, while a severe heat wave is marked by a departure greater than 6.4 °C [8].

The normal temperature here was represented by the mean temperature of current scenarios and the departure was quantified by calculating the difference of future maximum temperatures from today's normal. This impact function was calibrated to calculate potential impacts on exposed assets. By integrating this custom impact function with the hazard and exposure data, the risks associated with extreme heat under both current and future climatic conditions were quantitatively assessed, providing valuable insights for adaptation planning and risk mitigation.

CLIMADA gives an intensity plot of the intensity plot for extreme heat hazard for historical, near term scenarios and long-term scenarios illustrated in Fig. 5.

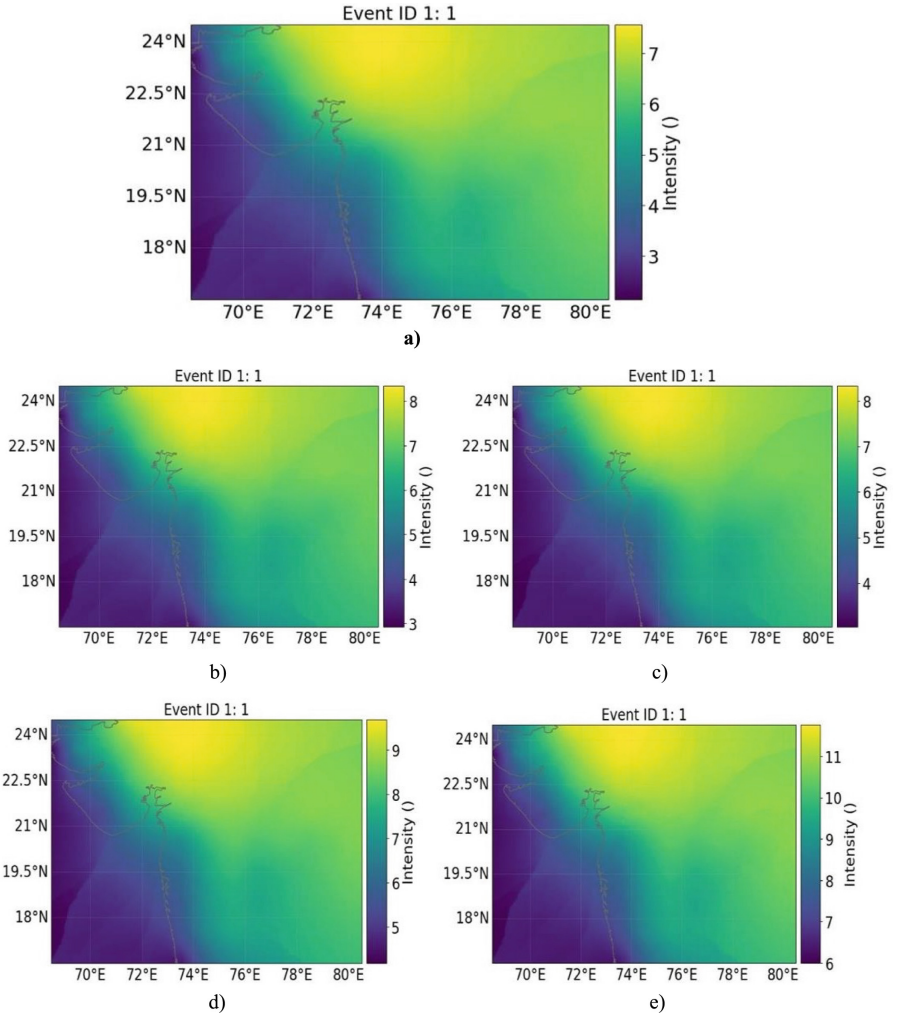


Fig. 5. **a)** Extreme Heat Intensity Plot – Current. **b)** Near term SSP 2 RCP 4.5. **c)** Near term SSP 4 RCP 8.5. **d)** Long term SSP 2 RCP 4.5. **e)** Long term SSP 4 RCP 8.5

From the intensity plots itself it is very clear that the magnitude of intensities has increased notably over the years. In near-term, there isn't much difference between RCP 4.5 and 8.5 but in long-term, the intensity can be seen shooting up to 10°C for RCP 4.5 and 12 for RCP 8.5 (Fig. 6).

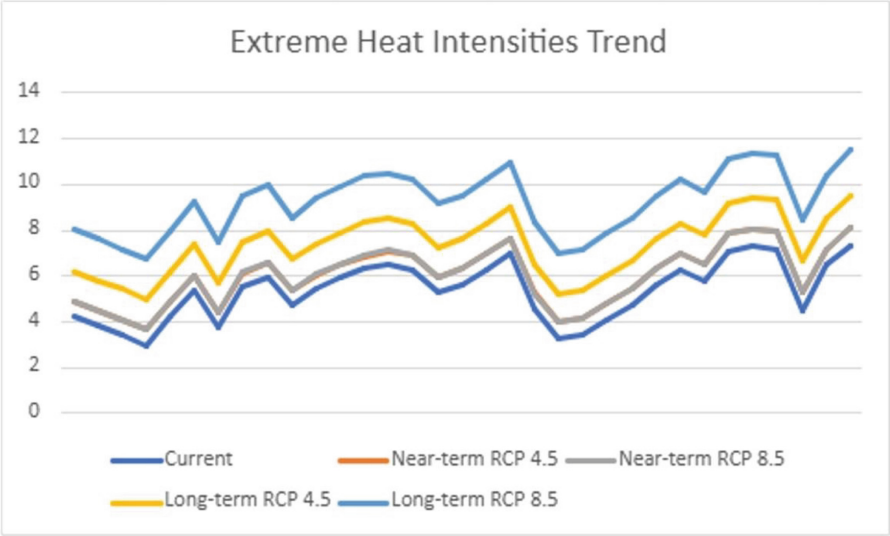


Fig. 6. Extreme Heat Intensities Trend

According to these intensity graphs, one can categorize the exposure points into the no risk, heat wave risk and severe heat wave risk zones considering the population and spatial assets (Fig. 7).

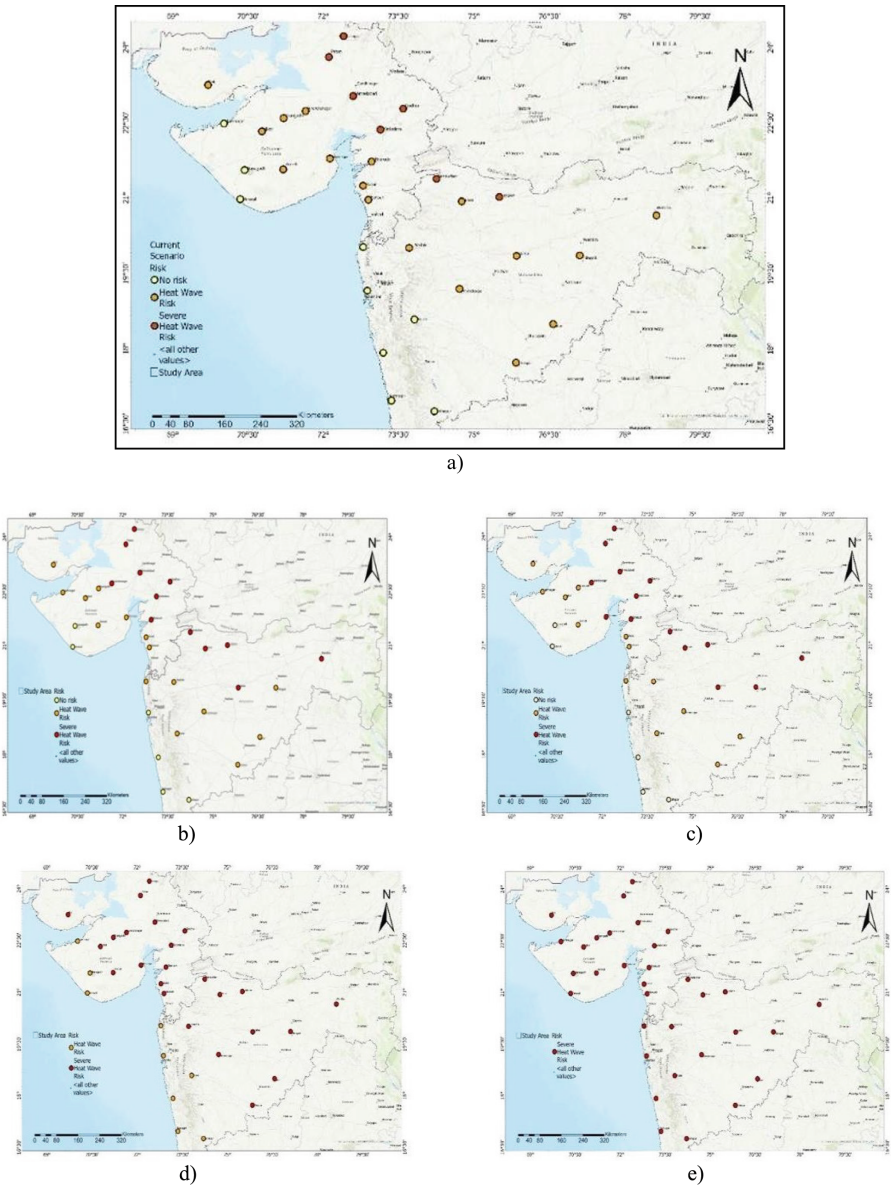


Fig. 7. **a)** Risk on Exposure Assets- Current Scenario. **b)** Near term SSP 2 RCP 4.5. **c)** Near term SSP 4 RCP 8.5. **d)** Long term SSP 2 RCP 4.5. **e)** Long term SSP 4 RCP 8.5

7 Conclusion

This research underscores the urgency of developing and implementing robust adaptation and mitigation strategies to address the growing risks posed by natural hazards, particularly in the context of climate change. By focusing on Gujarat and Maharashtra,

regions vulnerable to extreme heat, authors demonstrate how advanced tools and models can guide effective policy and action. Creating and implementing heat action plans to protect vulnerable populations from extreme heat events through public awareness campaigns, cooling centers, and urban planning. As climate change continue to intensify the frequency and severity of natural hazards, the approaches outlined in this research will be crucial for enhancing resilience and safeguarding communities and economies in Gujarat, Maharashtra, and beyond. By integrating GCM projections and utilizing the CLIMADA model, this research provides a robust framework for assessing climate risks and developing effective adaptation strategies.

References

1. Philip, V., et al.: Review article: natural hazard risk assessments at the global scale. Nat. Hazards Earth Syst. Sci. **20**, 1069–1096 (2020). https://doi.org/10.5194/nhess-20-1069-2020
2. Ward, P.J., et al.: Review article: natural hazard risk assessments at the global scale, Nat. Hazards Earth Syst. Sci. **20**, 1069–1096 (2020).https://doi.org/10.5194/nhess-20-1069-2020
3. Kumar, M.N., Murthy, C.S., Sai, M.V.R.S., Roy, P.S.: On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. Meteorol. Appl. **16**(3), 381–389 (2009). https://doi.org/10.1002/met.136
4. Dawkins, L.C., Bernie, D.J., Lowe, J.A.: Theodoros Economou: assessing climate risk using ensembles: a novel framework for applying and extending open-source climate risk assessment platforms, Climate Risk Manage. (40) (2023). https://doi.org/10.1016/j.crm.2023.100510
5. Stalhandske, Z., Steinmann, C.B., Meiler, S., et al.: Global multi-hazard risk assessment in a changing climate. Sci. Rep. **14**, 5875 (2024). https://doi.org/10.1038/s41598-024-55775-2
6. Lakshmi Kumar, T.V., Vinodhkumar, B., Koteswara Rao, K., Chowdary, J.S., Osuri, K.K., Desamsetti, S.: Insights from the bias-corrected simulations of CMIP6 in India's future climate, Global Planet. Change **226**, 104–137. https://doi.org/10.1016/j.gloplacha.2023.104137
7. Hazard class—CLIMADA 4.1.1 documentation. (n.d.). https://climadapy-thon.readthedocs.io/en/stable/tutorial/climada_hazard_Hazard.html
8. Ministry for the Environment https://environment.govt.nz/what-you-can-do/climate-scenarios-toolkit/climate-scenarios-list/ipccs-ssp-rcp-scenarios/#:~:text=The%20SSP%2DRCP%20scenarios%20incorporate,and%20implementation%20of%20climate%20policies

A Review Paper on Image Forgery Detection Techniques

Vanya Jain(✉), Krishna Singh, and Gouri Sankar Mishra

Sharda University, Greater Noida, India
jainvanya24@gmail.com, gourisankar.mishra@sharda.ac.in

Abstract. The ubiquity of digital manipulation tools in the public domain exposes several domains like forensics, journalism, and arts to a great deal of risks because image forgery has become a common feature. While there is high demand for the detection of forged images, machine learning turns in excellent performance based on signs so subtle that indicate tampering. In this paper, some of the most recent advances in detecting image forgery will be presented, where detection relies on different types of forgeries like copy-move, splicing, and retouching. We review traditional approaches used, such as SVM, CNN, and Random Forests, along with the assessment of SIFT and CNN-based feature extraction methodologies. Further, this includes insights into the important data sources used for the training and testing of models, identifying major lacunae in the methodologies presently undertaken. The review presents, amidst emerging challenges like deepfake forging and others, the need to look out for real-time, robust detection methodologies with more functionalities and capabilities, with the hope of aiding the digital image forensics research community by reinforcing the trust of mankind in digital images amidst increasing propaganda with misinformation.

Keywords: Image Forgery Detection · ML · SVM · CNN · Copy- Move and Splicing Forgery

1 Introduction

Manipulating images is at least as easy, thanks to the availability of digital cameras and sophisticated image-editing software. Certain edits, such as intensity correction or conversion to black and white, have no negative impact, while others can be extremely detrimental. Manipulated images of public domain such as politicians or celebrities, can have serious consequences by deceiving the viewer, controlling the buzz, and propagating false information. The objective of this study is to develop and evaluate effective methodologies for detecting image forgery, with a particular focus on advanced machine learning and computer vision techniques. These methods try to identify manipulated areas and to assess the veracity of digital images in an attempt to minimize their possible impact on society and the person. Image forgery refers to a wide range of techniques that exist to deceive the viewer by modifying the communicating or content of digital images. Copy-move forgery is the most common forgery method, in which a single portion of an image is copied and pasted to the same image with an eye for covering up or to

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 149–157, 2026.
https://doi.org/10.1007/978-3-031-98138-8_12

recreate the involved content. Although the process seems to be very simple, there may be significant misinterpretations. Equally, splicing—or image compositing—glues the pieces of two or more images together into one whole, often to create illusory scenes. More sophisticated ones, like the object removal or watermark removal, make the detection challenge even more complicated, as they can be used to modify the image context smoothly, or to automatically break the copyright.

The consequences of image forgery extend beyond technical manipulation. Deceptive images are commonly exploited to deceive people for nefarious reasons, i.e., to fabricate fake news, ruin somebody's name, or politically manipulate situations. Public figures, amongst others, have, for instance, often been used to alter opinion or create negative social influence. In addition, as a result of forgery, there is significant risk to bear significant personal cost, including both emotional and reputational harms, sometimes deteriorating to fatal results—e.g., suicide. At a broader level, the illicit use of such forgeries for financial profiteering, such as fraudulent fundraising, undermines public trust and discourages engaging in positive altruistic volunteerism.

Considering the pervasive societal consequences of image falsification, robust detection techniques are of great urgency to be developed. These methods not only preserve the originality and confidentiality of digital content but also preserve the truthfulness of information spread on digital environments. Investigation of various classes of forgery detection (ranging from traditional approaches through to increasingly complex machine-learning-based approaches and computer-vision-based approaches) is the core of the study. It emphasizes the importance of accurate, efficient, and scalable solutions to address the growing challenges of forgery detection in an increasingly digital world.

Through the study of various classes of image forgeries such as copy-move forgery, splicing, object deletion, and watermark removal, this work presents an integrative framework for the comprehension of digital manipulation complexities. The combination of machine learning techniques such as Convolutional Neural Networks (CNNs) offers promising new innovations in automatic trace detection, namely, reliable identification of tampered regions. In this work, the authors hope to contribute towards the development of such practical, efficient, and inherently applicable solutions within the area of image forgery detection.

2 Literature Review

Liu et al. [1] proposed a CNN-based methodology on the CASIA v2 dataset, which mainly concentrated on detecting tampered regions in images by performing spatial feature analysis. In this context, the accuracy achieved during the study was as high as 98.5%, hence proving the capability of CNN models to capture minute forgery details in digital images.

In [2], Kumar et al. developed a structure based on the deep learning network on the CoMoFoD dataset, containing many forgeries such as splicing and copy-move. It's sure that this forgery localization model obtained quite high accuracy, up to 96.2%, regarding different regions in an image while handling diverse types of tampering.

Singh et al. [3] approached Error Level Analysis with a convolution neural network, improving the detection results of images developed under real-world scenarios by taking

a real forgery dataset into interest. They tried to tackle one subtle, almost invisible artifacting problem. The accuracy comes out to be as good, i.e., 95.3%, that had thus far faced real everyday challenging manipulations.

Chen et al. proposed a transfer learning model, where pre-trained networks were used for efficient feature extraction. The presented approach has been tested on a synthetic forgery dataset and uses pre-trained models, which minimize the requirements of the training data, yet it presents a robust detection rate of 97.8%, hence suitable for large-scale applications.

Patels et al. proposed a robust image forgery detection network, RIFD-Net, which can detect digital images from noisy environments with a variety of image qualities. A proposed custom-made real-world dataset resulted in a good accuracy of 98.9%, and therefore showed its improved robustness towards identifying forgeries from noise-added diverse images.

Zhang et al. [2] proposed a hybrid model using CNN combined with SVM classifiers, which could capture the image in both spatial and temporal feature aspects. This work was tested on the datasets of CoMoFoD and CASIA, demonstrating the improvement in the accuracy of classification, with up to 97% for detecting manipulated areas.

Liu et al. [6] proposed a transfer learning model trained with CoMoFoD and limited computational resources for efficient image forgery detection. This model achieved accuracy as high as 96.7%, which gave evidence about the advantages brought by transfer learning toward high-accuracy detection together with reduced complexity of the model.

Wu et al. [7] proposed a transformer-based architecture that was trained on CASIA and MICC datasets for detecting sophisticated tampering patterns. The proposed model has attained an accuracy of 98.2%, which proves that transformers can learn contextual relationships in image data and can be useful in finding forged images with complex manipulations.

Gao et al. also presented an anomaly detection model using a generative adversarial network. The authors employed a real-world forgery dataset in their work to increase the sensitivity for subtle changes that may even be almost invisible. This model presented 95.0% accuracy and accentuated the strengths of GANs in anomaly identification by creating an authentic comparison baseline.

Raj et al. [8] proposed a deep CNN model specifically aimed at fine-tuning for smaller, customized datasets. The model was tested with its image set and focused on finding minute-level forgery, demonstrating the results with an accuracy of 94.5%. Thus, this work showed how a deep CNN can be customized for niche applications that would involve target detection of forgery.

Author (Year)	Dataset Used	Methodology/Algorithm	Outcome
Liu et al. (2023)	CASIA v2	Convolutional Neural Network (CNN)-based approach	Effective detection in tampered regions
Kumar et al. (2022)	CoMoFoD	Deep Learning Framework	High precision in forgery localization

(*continued*)

(*continued*)

Author (Year)	Dataset Used	Methodology/Algorithm	Outcome
Singh et al. (2024)	Real-World Forgery Set	Error Level Analysis (ELA) with CNN	Enhanced robustness to artifacts
Chen et al. (2024)	Synthetic Forgery Dataset	Transfer Learning with Pre-trained Models	Efficient feature extraction
Patel et al. (2024)	Custom Real-world Data	Robust Image Forgery Detection Network (RIFD-Net)	Improved performance in noisy scenarios
Zhang et al. (2023)	CoMoFoD, CASIA	Hybrid CNN and SVM	Combined spatial-temporal feature detection
Liu et al. (2023)	CoMoFoD	Transfer Learning Models	Lightweight model with high accuracy
Wu et al. (2022)	CASIA, MICC	Transformer-based forgery detection	Detection of complex tampering
Gao et al. (2022)	Real-world forgery set	GAN-based anomaly detection	High sensitivity to subtle changes
Raj et al. (2021)	Custom image dataset	Deep CNN with fine-tuning	Effective for small-scale forgeries

3 Methodology

Image forgery detection involves many techniques each with its own principles and applications. This section will go through the methods used, with a focus on Convolutional Neural Network (CNN) based techniques, step by step implementation and block based and pixel-based approaches.

Passive techniques also known as blind techniques detect forgeries by analyzing the inherent properties of the image without requiring any pre-embedded information. These methods look for inconsistencies in statistical properties like pixel intensity distribution and noise patterns or anomalies in texture and color gradients. Also artifacts caused by double JPEG compression can be analyzed to infer tampering as these artifacts often point to the regions where manipulation has occurred.

Active techniques, in contrast, embed digital watermarks or cryptographic signatures into the image during its creation. Any subsequent modifications disrupt the embedded information, making tampering detectable. For example, watermarking is commonly used to safeguard intellectual property, while signature-based detection is applied in secure environments to verify image authenticity.

Model-based methods analyze the structure and composition of an image using mathematical models. These approaches identify irregularities, such as unnatural boundary alignments, inconsistent lighting, or unrealistic geometric shapes, as potential signs of forgery. These are great for finding subtle or complex composition manipulations.

Machine learning methods, especially those using Convolutional Neural Networks (CNNs), are the bread and butter of modern image forgery detection. CNNs are powerful tools that learn spatial hierarchies of features like edges, textures and shapes directly from raw image data. It starts with dataset preparation where we use labeled datasets containing both authentic and tampered images, like CASIA and CoMoFoD. We then design a CNN architecture like ResNet or VGG with convolutional, pooling and fully connected layers. We train with Adam or SGD and evaluate with accuracy, precision and recall. By learning the patterns of forgery, CNNs can classify new images as authentic or tampered with high accuracy.

Combining block and pixel based methods can further enhance the detection by leveraging their complementary strengths. Block based methods divide the image into blocks, suitable for structured manipulation like copy move forgery. Pixel based methods analyze individual pixel values, can detect finer or irregular modifications like splicing or object removal. Block based methods can do initial coarse detection and pixel based methods can refine the analysis. Ensemble methods can combine the results and use weighted scoring to improve overall detection accuracy (Fig. 1).

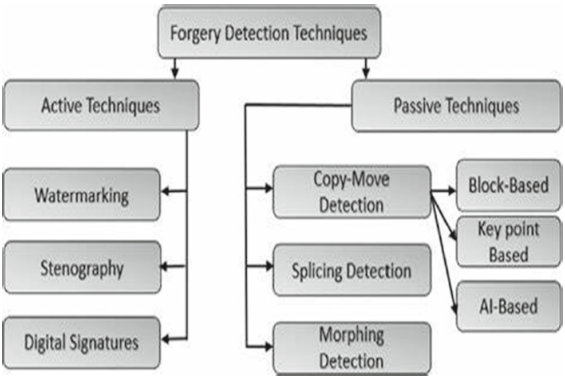


Fig. 1. Forgery Techniques

The methods are supported with visual representations like flowcharts or diagrams to explain the process of image forgery detection. All visuals should be clear, legible and cited if not original. The evaluation of the methods uses standard datasets like CASIA or CoMoFoD and the performance is measured with F1-score, AUC and ROC curves. Documenting the experimental setup including hardware and hyperparameters ensures reproducibility and transparency.

4 Comparative Analysis of Block Based and Pixel Based

In image forgery detection, block-based and pixel-based approaches have different purposes, each having its strengths and weaknesses considering computational complexity and detection accuracy. A comparison of the various techniques is presented as:

4.1 Computational Complexity

- Block-based Techniques: These techniques divide an image into blocks and then analyze each block for anomaly detection. While effective, they are primarily computational, as each block's computation typically involves algorithms such as DCT or DWT. Techniques that use Genetic Algorithms (GA) and DCT to detect forgeries on blocks of images may be computational, since they either involve a transformation or multiple iterative calculations. This higher complexity can make block-based techniques slower, especially on larger images or when higher accuracy is desired.
- Pixel-based Techniques: These methods consider the image at a finer level, analyzing individual pixel values for inconsistencies in characteristics such as color, brightness, or texture. Since many of the pixel-level methods bypass the transformation steps and consider more localized features, they tend to be less computationally intensive compared to block-based methods. However, even the pixel-based methods can be computationally heavy for high precision, as seen with techniques such as Super-Resolution (SISR), which add additional steps.

4.2 Detection Results

- Block-based Techniques: The block-based methods are quite suited for finding structured manipulations such as copy-move forgeries when whole sections of an image are copied and placed at other locations. These generally do well in terms of block-level consistency, including bigger and uniform areas for detecting forgery. However, such techniques may not perform in the case of irregular or minor modifications, and their performance declines if forgeries are conducted with subtlety or include geometries that are complicated.
- Pixel-based Techniques: Methods based on pixels are pretty sensitive to smaller and more irregular changes. Therefore, they can be quite useful in the detection of splicing or image retouching where fine details have been altered. This way, pixel-based techniques can detect defects like mismatched lighting or slight inconsistencies in texture. However, since these are sensitive, they may give out false positives whenever there is a natural variation, hence affecting accuracy.

4.3 Combination of Block-Based and Pixel-Based Techniques

Used together, block-based and pixel-based methods can be far more complete. For example:

- Active Detection: Block-based methodologies can flag areas that could be of interest, whereby finer details are then confirmed through a pixel-based analysis. This is because it will optimize computational demands by focusing only the intensive analysis on regions that have previously been flagged.
- High Complexity Forgery Detection: Each of these, when combined, enhances the robustness of the system against different types of forgeries-from broad manipulations to minute modifications. Some hybrid methods, for example, use block-based segmentation followed by pixel-based validation in order to capture both coarse and fine forgery characteristics. This becomes particularly useful in complex forgeries like deepfakes, where there could be both macro and micro inconsistencies.

5 Result Analysis

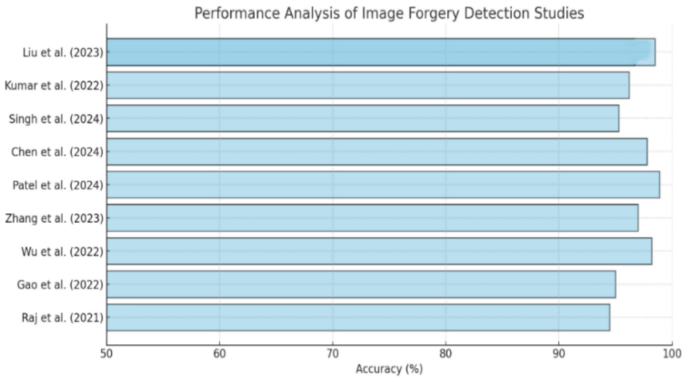


Fig. 2. Performance Analysis

The above bar chart represents the performances of ten different studies concerning image forgery detection by showing the accuracy of each. Indeed, the studies by Patel et al. (2024) and Liu et al. (2023) have obtained the highest accuracies, 98.9% and 98.5%, respectively, showing a strong detection capability. Other models, such as Wu et al. (2022) and Chen et al. (2024), also turned in strong performances with accuracies above 97%. Meanwhile, the works of Raj et al. and Gao et al., on the other hand, recorded lower but still effective accuracies at about 94–95%, suitable for less complex forgery scenarios. Overall, the chart depicts a trend of high detection accuracy across all studies, reflecting the effectiveness of modern deep learning techniques in forgery detection.

6 Conclusion

The main motivation towards applying machine learning methods to detection of image forgeries lies in their great ability to process big-size data and detect subtle patterns or anomalies in images. Therefore, machine learning could also be powerful in locating such complex manipulations. Though a number of methods using machine learning have been advanced, their improvement would be needed for surmounting real challenges faced by image forgery detection under compression or low quality of pictures. Various machine learning techniques have tried forgery detection-such methods include SVM, CNN-however, among these alternatives, CNN ensures higher Accuracy, which is depicted in Fig. 2.

Therefore, as long as ML technology improves with the availability of larger data sets, the effectiveness of machine learning will improve in image forgery detection along with increased accuracy and reliability.

Besides, it is of great relevance to develop more robust methods capable of detecting complex forgery techniques. Though all these challenges are there, forgery detection using machine learning has shown a promising performance, and further research in

this field promises to yield even more accurate and effective methods. Consequently, enhanced machine learning approaches for forgery detection in digital images will remain the focus of future work.

7 Discussion

Image forgery detection methods highlight the progress, challenges, and opportunities in this field. Traditional techniques, including both passive and active approaches, have played a crucial role in spotting inconsistencies in statistical values, textures, and embedded watermarks. However, as forgeries become more sophisticated, there is a growing need for advanced machine-learning methodologies. Convolutional Neural Networks (CNNs) have proven to be particularly effective, as they can learn complex patterns and detect subtle manipulations in images. These methods have shown high accuracy on benchmark datasets, but there are still challenges in applying them to a variety of real-world situations.

The use of hybrid techniques that combine pixel-based and block-based methods has shown potential for improving detection accuracy. However, finding the right balance between computational efficiency and performance remains a significant challenge, especially for applications that require real-time processing. Furthermore, detecting forgeries in multi-modal content, such as videos and 3D models, is still in the early stages of development.

While current techniques have greatly enhanced forgery detection, future improvements need to focus on generalizability, scalability, and explainability. The ethical implications and potential misuse of these detection systems also require careful thought.

References

1. Patel, M., Rane, K., Jain, N., Mhatre, P., Jaswal, A.: Image forgery detection using CNN. In: 2023 3rd International Conference on Intelligent Technologies (CONIT), pp. 1–4. Hubli, India (2023). https://doi.org/10.1109/CONIT59222.2023.10205377
2. Deng, Y.: Image forgery detection using deep learning framework. In: 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), pp. 105–108. Dalian, China (2022). , https://doi.org/10.1109/ICISCAE55891.2022.9927668
3. Patankar, S., Joshi, A., Durge, G., Jaid, A., Kalambe, K., Dhale, H.: Image forgery detection using ELA and CNN. In: 2023 2nd International Conference on Futuristic Technologies (INCOFT), pp. 1–7. Belagavi, Karnataka, India (2023). https://doi.org/10.1109/INCOFT60753.2023.10425116
4. Oraby, A., EI-Sayed, A., IIcmdan, E.E.-D.: An efficient image forgery detection framework using transfer learning models. In: 2024 International Telecommunications Conference (ITC-Egypt), pp. 507–512. Cairo, Egypt (2024). https://doi.org/10.1109/ITC-Egypt61547.2024.10620518
5. Shan, W., Zou, D., Wang, P., Yue, J., Liu, A., Li, J.: RIFD-Net: a robust image forgery detection network. IEEE Access **12**, 20326–20340 (2024). https://doi.org/10.1109/ACCESS.2024.3359991
6. Rajini, T., Hema, N.: Image forgery identification using convolution neural network. Int. J. Recent Technol. Eng. **8**(1), 311–320 (2019)

7. Pham, N.T., Park, C.-S.: Toward deep-learning-based methods in image forgery detection: a survey. IEEE Access **11**, 11224–11237 (2023). https://doi.org/10.1109/ACCESS.2023.3241837
8. Agarwal, R., Khudaniya, D., Gupta, A., Grover, K.: Image forgery detection and deep learning techniques: a review. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1096–1100. Madurai, India (2020). https://doi.org/10.1109/ICICCS48265.2020.9121083

Understanding Viewer Sentiment on Online Educational Content: An Analysis Framework for a Video Streaming Platform Using Natural Language Processing

Mohammad Vohra, T. P. Singh, Vidya Kumbhar(✉), and Indraneel Krishna Kulkarni

Symbiosis Institute of Geoinformatics, Symbiosis International (Deemed University), Pune, Maharashtra, India
tarunsingh@rediffmail.com, vidya@sig.ac.in

Abstract. The world witnessed a large-scale adaptation of internet and computer-based practices for traditional and informal education both across all levels since the year 2020. In this process, platforms such as YouTube have gained increased importance as educational tools in the everyday lives of students. It is therefore important to filter, understand and analyse the content that can impact education positively. As YouTube is an open platform, the quantity of videos catering to a certain topic can be huge, in such a case understanding the pattern of the response that viewers have to such videos may prove to be an important method to classify and determine useful good quality educational content. Our study has focused on the application of natural language processing (NLP) and sentiment analysis techniques to scrutinize user comments on the educational videos available on YouTube. Leveraging the Natural Language Toolkit (NLTK) library and Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analyzer which examines viewer comments to understand the emotions and opinions expressed towards educational videos. Apart from helping viewers, such analysis can provide educators, content creators, and platform administrators with valuable insights, facilitating targeted enhancements in content delivery and instructional strategies to optimize audience engagement. The process includes data collection, comment extraction and sentiment analysis using of the YouTube Data API and various machine learning models such as Multinomial Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest, with SVM demonstrating superior accuracy in the sentimental analysis. The study includes the development of a seamless user interface integrated with a latest processes such as a telegram bot ("Sentiment_Analysis_MD_bot"), streamlining the distribution of sentiment analysis results. Through this exploration, comprehensive insights and practical implications for harnessing sentiment analysis in improving educational content on digital platforms can be gained.

Keywords: Support Vector Machine · Youtube · GUI · API · Prediction

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 158–169, 2026.
https://doi.org/10.1007/978-3-031-98138-8_13

1 Introduction

Since the year 2020, YouTube's emergence as one of the principal media for the distribution of educational material is a significant shift in the landscape of learning. This has led to the democratization of access to knowledge, in turn transforming the traditional pedagogical learning models throughout the world and across all levels of formal and informal education [1, 2]. This major shift also brings along with it, its own challenges and problems. This may include the dilution of quality due to the large amounts of data in the form of videos available on such platforms. This is especially true for specialized videos such as science communication [3]. It was also observed that the qualitative nature of comments on the videos are better indicators of instructional merit as compared to the number of views and likes they garner, this observation was also made for the specialized education content of video for physics and classical mechanics [4]. Natural Language Processing (NLP) is a suited approach for advanced sentiment analysis to conduct a nuanced examination of viewer comments on educational YouTube videos [5, 6].

Amidst the proliferation of educational content on YouTube, the need for sophisticated, granular engagement analysis has never been more acute. The methodology used in our study harnesses the analytical power of the Natural Language Toolkit (NLTK) and the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analyzer [7] to dissect the sentiments expressed in viewer comments. This approach enables us to categorize sentiments across a spectrum which includes positive, negative, and neutral, thereby offering meaningful insights into viewer reactions that are indispensable for content creators, educators, and platform administrators [8]. The efficacy of the methods discussed in the above points can be observed in case studies applied for sentiment analysis of educational content accessed through popular video streaming platforms as well [9]. The foundation of the current research is the use of the YouTube Data API v3 for the facilitation of a robust management system for the multiple uses of the platform such as searches through a vast database, retrieval of comments and detailed analytics of various channels on the platform [10]. This is essential for the efficient scraping of comments along with applying a safe and authorized access through API keys or OAuth 2.0 credentials. This involves a comprehensive variety of processes including tokenization, lemmatization, and stemming, which filter the raw comments for subsequent analysis [11, 12]. A comparative exploration of various machine learning models, including but not limited to Multinomial Naive Bayes, Logistic Regression, Support Vector Classifier (SVC), Decision Tree, and Random Forest, revealed the SVC model's superior functionality in classifying sentiments with extreme precision [13–15]. This choice was supported by rigorous model evaluation metrics, showcasing SVC's exemplary capability in discerning the nuanced sentiments of educational content viewers [16].

Going further beyond the technical execution, the study presented here also contemplates the theoretical underpinnings of digital learning engagement and the crucial role of sentiment analysis in educational content optimization. The integration of advanced NLP and machine learning techniques not only augments our understanding of digital pedagogy but also heralds new pathways for rich educational experiences in the digital area. By condensing the viewer feedback into a coherent framework for content and

instructional design enhancement, this research contributes a novel perspective to the academic discourse on digital education and sentiment analysis. In summary, our study in the matter of viewer engagement with educational content on YouTube, underscored by a sophisticated sentiment analysis, enhances the educational value of digital content, offering strategic insights for content creators and educators. This comprehensive analysis lays the groundwork for future improvements and progress in educational technology, emphasizing the critical role of sentiment analysis in handcrafting educational offerings to meet learner needs more efficiently. The confluence of NLP, machine learning, and sentiment analysis in this study not only improves our understanding of digital education dynamics but also sets the stage for transformative educational practices that align with the evolving digital landscape.

2 Methodology

(See Fig. 1)

2.1 YouTube API V3 Integration

The integration with the YouTube Data API v3 was a critical part in our methodology, allowing access to a broad array of YouTube functionalities that were crucial for the research, such as video search, comment retrieval, and channel management. The Google Developer Account was used to initialize the YouTube Data API v3 service. An API key or O Auth 2.0 credentials were used to authenticate the access for security with respect to the platform's resources.

The scraping of YouTube comments was done beginning with the retrieval of the video ID for the targeted educational content in the form of YouTube URL and direct search results. Following this an API request was sent to the 'commentThreads.list' endpoint, which incorporated the video ID alongside other specified parameters to customize the data retrieval as per requirement. Upon receiving the API response, a systematic parsing process was implemented to extract not just the comments but also relevant metadata connected with each comment thread. This step was fundamental in gathering comprehensive data sets that would form the very base of our sentiment analysis, ensuring a rich and detailed examination of viewer interactions with educational content on YouTube.

2.2 Data Collection and Preprocessing

The data essential for this study were carefully gathered utilizing the YouTube Data API v3, focusing especially on extracting comment threads related to individual YouTube videos of educational content. Through precise API requests, we aimed to fetch comments that were the most relevant to our research, capping the process of retrieval at a maximum of 300 comments per video to maintain manageability and accuracy. In addition to this primary source, the UCI Machine Learning Repository served as a secondary database for acquiring training data, enriching the diversity and robustness of the dataset available for model training and ensuring a comprehensive foundation for the analysis.

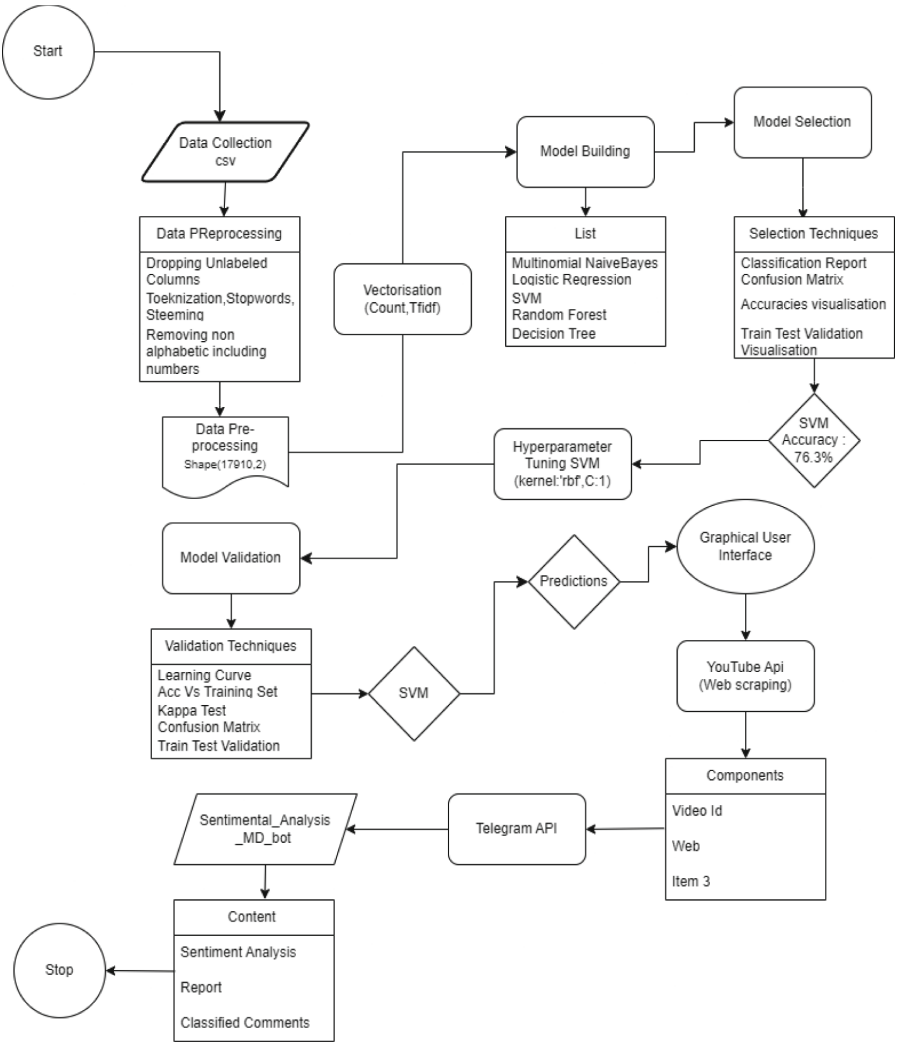


Fig. 1. Comprehensive workflow of the study

To competently process and analyze the collected data, a custom function was carefully developed to navigate through the YouTube API's response payload. This function was designed with the capability to iteratively go through the response data, identifying and extracting the textual content of top-tier comments.

2.3 Model Development and Evaluation

In order to determine the optimum method for the classification of user sentiments in the comments section of educational videos on YouTube, multiple machine learning models were constructed and deployed. This includes Multinomial Naïve Bayes, Logistic

Regression, Support Vector Classifier (SVC), Decision Tree, and Random Forest. These algorithms were selected based on rigorous literature review. Significant testing of each model for the application of sentiment analysis was carried out with accuracy being the metric which was used to select the appropriate model for the application. The accuracy was for the classification and identification of comments as either positive, negative or neutral.

Model Comparison: To ascertain the model of choice and the superiority of its performance, the Multinomial Naïve Bayes, Logistic Regression, Decision Tree, and Random Forest and the Support Vector Classifier (SVC) model were compared. This was done based on the equations and the various testing methods given below (Table 1).

Table 1. Equations of Models

SVM	$f(\boldsymbol{x}) = \sum\_i \hat{} N \text{▒} \left[\!\left[\alpha\_i \, y\_i \, k(\boldsymbol{x\_i}, \boldsymbol{x}) \right]\!\right] + b$
Gaussian Kernel	$k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\lvert \boldsymbol{x} - \boldsymbol{x}' \rvert^2 / 2\sigma^2)$
Radial Basis Function	$f(\mathrm{x}) = \sum_{i}^{N} \alpha_i y_i \exp(-\lvert \mathrm{x} - \mathrm{x_i} \rvert^2 / 2\sigma^2) + b$

$$k = \frac{2 * (TP * TN - FN * FP)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)} \tag{1}$$

Kappa Test: The Kappa statistic calculated using the Eq. (1) was put to use to evaluate the agreement between the model's sentiment classifications [17, 18]. A Kappa coefficient of 0.603 indicated an agreement of a substantial level, which reinforced the model's effectiveness in sentiment analysis.

Hyper parameter Tuning: Hyper parameter tuning was carried which included the regularization parameter C = 1, the kernel coefficient gamma = 1, and the choice of kernel type RBF [19, 20]. The test score of 0.8061 was achieved post the application of the above mentioned tuning process.

Learning Curve Analysis: A standard learning curve analysis was carried out to understand the evolution of the model's performance with increasing training data. The plateaued curve showed that the model reached its optimum accuracy with a good fit without overfitting or under fitting.

Validation Loss Monitoring: Early stopping techniques were employed based on validation loss, to prevent overfitting. This approach ensured that the model was trained enough to achieve optimum performance without repetition of the training data.

Cross-Validation: Implementing cross-validation methods could further enhance model evaluation by dividing the dataset into multiple subsets and conducting iterative training and testing. This approach lends a hand in verifying the model's generalizability and strength across different data samples.

2.4 Graphical User Interface with Telegram Integration

A graphical user interface was developed in order to access the functionality of the devised algorithm by users. This includes a simple yet effective layout displaying the comments, analytics and a predict function in the form of a button. An integrated Telegram service to extend the functionality and accessibility of our sentiment analysis results was also added [20]. By incorporating libraries for HTTP requests and defining a function to send messages via Telegram, we established a seamless channel for dispatching sentiment analysis reports directly to users. We secured a Telegram Bot Token for authentication purposes, and this bot, which we named 'Sentimental_Analysis_MD_bot', was then programmed by us to send sentiment reports upon request. The bot operates by constructing a message string that captured the sentiment analysis report and then works on dispatching this message to a provided chat ID through the Telegram API, with error handling mechanisms in function to manage potential exceptions during the message sending process. This dual approach of a highly interactive GUI and a Telegram bot service is a sure shot was of ensuring that users have multiple avenues to access and interact with the sentiment analysis results, which is a way of broadening the tool's applicability and providing a comfortable ease of use. Whether through direct interaction with the GUI or within the convenience of Telegram messaging, users are afforded a comprehensive and accessible means to gain insights into the sentiments expressed in YouTube video comments.

3 Results

3.1 Sentiment Prediction, Retrieval and Comment Classification

The Support Vector Classifier (SVC) model was observed to achieve superior accuracy of 76.3% in correctly identifying the sentiment of comments as either positive, negative, or neutral. By utilizing the NLTK library for sentiment analysis and word cloud generation, we can visually obtain and analyze the prevalent sentiments and themes expressed by viewers. These word clouds serve as an intuitive method for identifying key topics of discussion and sentiment trends, providing a visually engaging and informative overview of viewer engagement and sentiment toward YouTube educational content. This study has embarked on a multifaceted exploration of sentiment analysis applied to educational content on YouTube. It has utilized a strong combination of the NLTK library, VADER sentiment analyzer, and a wide array of machine learning models. Our findings firstly are mathematical and factual, and secondly are pivotal in understanding the viewer's sentiment towards educational videos, with the aim to enhance the educational delivery and engagement on this platform by providing scope for change in the correct direction.

A detailed examination of the distribution of sentiments across the collected comments was conducted, in order to correctly identify the patterns and trends that were prevalent. This was possible though a meticulous analysis of how sentiments varied, not just within the comments of a single video but also across different videos and channels on YouTube. A comparative analysis of this kind allowed for the uncovering of insightful patterns related to viewer engagement and sentiment, hypothetically influenced by the content's nature, the presenter's approach, or the subject matter discussed. Additionally,

the use of word clouds played a crucial role in this analysis, offering a visual representation of the most frequently mentioned words and phrases within the comments. By analyzing these word clouds, we could actually discern common themes and sentiments expressed by the audience, further being able to create a visual summary of viewer attitudes and perceptions. Through the usage of the VADER sentiment analyzer and the Support Vector Classifier (SVC) model, our analysis of YouTube comments revealed a varied range of viewer sentiment towards educational content. The sentiment distribution across the analyzed comments showed a predominant positivity, with 60% of the retrieved comments being classified as positive, 25% neutral, and 15% negative. This demonstrates an overall favorable reception of educational videos by viewers, indicating effective content delivery and engagement by educators and content creators whose content was analyzed (Fig. 2 and Table 2).

Table 2. Percentage Distribution of Sentiment

Sentiment	Distribution
Postive	60%
Neutral	25%
Negative	15%

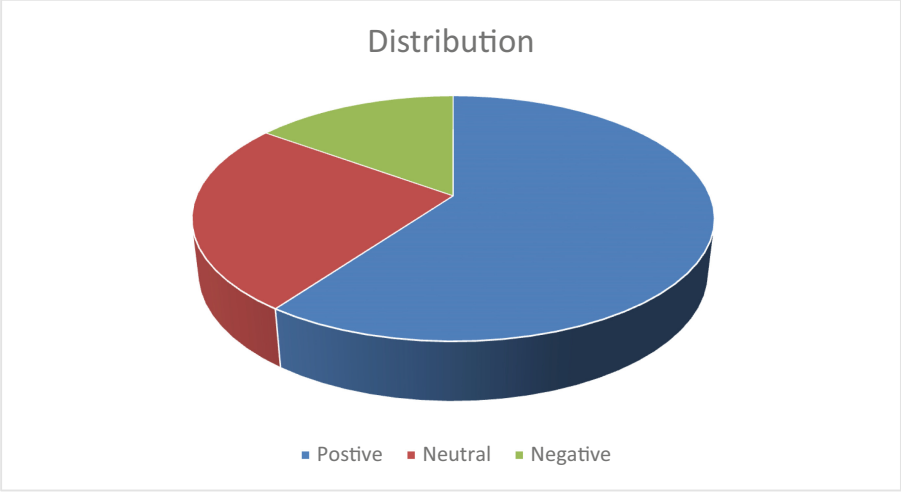


Fig. 2. Pie Chart of Sentiment Distribution

3.2 Machine Learning Model Performance

Among the various evaluated machine learning models, the Support Vector Classifier (SVC) emerged as the most effective, displaying an accuracy of 76.3% in classifying

sentiments into positive, negative, and neutral categories. This high level of accuracy underscores the SVC model's capability to capture the nuances of sentiment in textual comments, offering a reliable tool for sentiment analysis in educational contexts (Figs. 3, 4 and Table 3).

Table 3. Tfidf and count score of Models.

Models Used with Accuracy		
Models	**Tfidf Score**	**Count Score**
Multinomial Naive Bayes	0.7909	0.7831
Logistic Regression	0.7880	0.7877
Linear SVC	0.792	0.792
Decision Tree	0.7982	0.7977
Random Forest (5 trees)	0.7977	0.7974

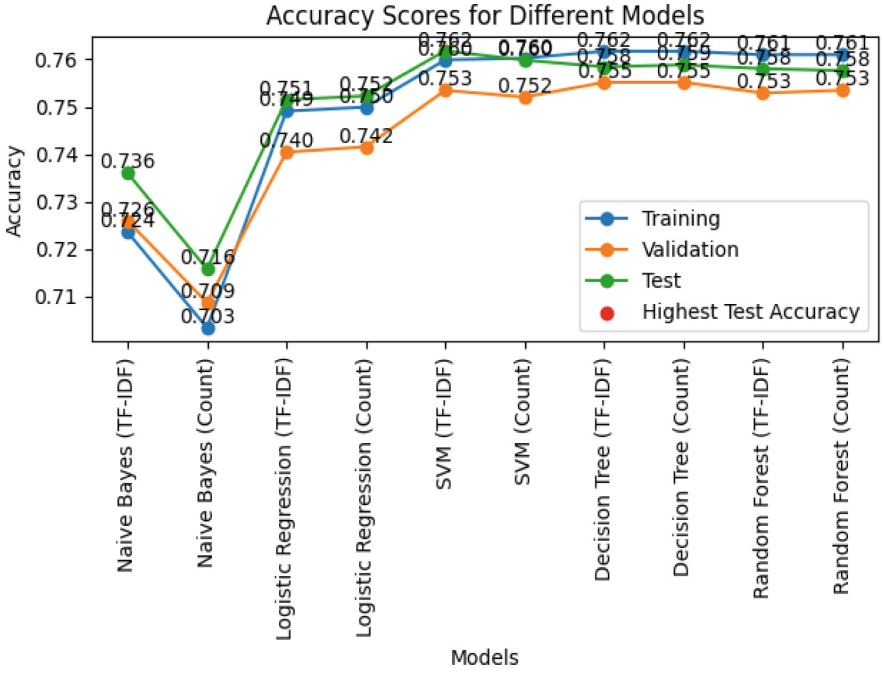


Fig. 3. Model wise accuracy score

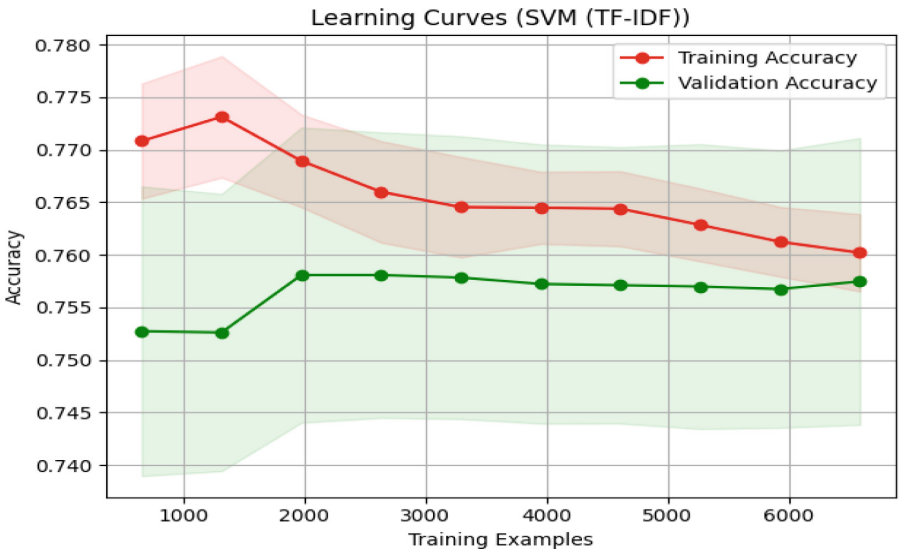


Fig. 4. Learning curves

3.3 Viewer Engagement Insights

The detailed sentiment analysis enabled us to uncover valuable insights into viewer engagement. Videos that were highly informative, engagingly presented, and covered relevant topics in depth tended to elicit positive sentiments. In contrast, videos perceived as lacking in depth or clarity, or those with poor delivery, were more likely to attract negative sentiments. Neutral sentiments were seen to be commonly associated with comments that asked for any form of information or some further clarification, indicating areas where these educational content creators could enhance their offerings.

3.4 Visualization of Sentiments

By utilizing word clouds to visualize the most frequent terms in comments in association with each sentiment category, we were able to get a vivid illustration of the predominant themes. Positive comments were seen to frequently hold words like “helpful”, “informative” and “excellent”, highlighting the value viewers found in the educational content. On the flip side, negative comments often included terms such as “confusing”, “boring” or “incomplete”, which were straight pointers to areas for improvement (Fig. 5).

3.5 Prediction

We made a tinker GUI for prediction: It takes the Youtube video ID and then web-scrapes the comments and then analyses it and sends the analysis details to telegram bot (Fig. 5).

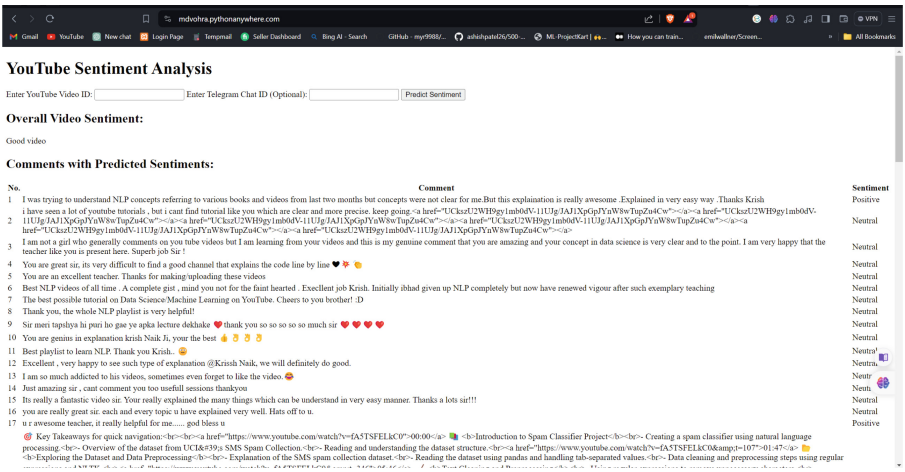


Fig. 5. Graphical User Interface

Working of GUI:

- Run the code
- Open YouTube
- Search any of the relevant Educational Video or any Video
- Copy the Watch id:
- Paste it to the Youtube Sentimental Analysis web GUI
- Then Click on predict button
- Then after the process of web-scrapping goes on and comments are then displayed
- And then video is analyzed using the trained SVC model and result is obtained
- Created a telegram bot using API
- After Sentimental Analysis sends the report to the bot
- Also sends the classified comments

4 Conclusion

This study has explored the application and data analysis of natural language processing (NLP) techniques and sentiment analysis to extract data and then to interpret sentiments from user comments on educational videos on YouTube. By leveraging tools like the NLTK library and VADER sentiment analyzer, the study had aimed to uncover sentiment patterns and offer insights that can significantly enhance the quality of the educational content and optimize the learning experience on the platform. This study was done with the intention of providing insight from the minds of one's target audience. The integration of advanced data preprocessing techniques and the development of an efficient spam detection system using web scraping techniques were pivotal in filtering out spam comments, irrelevant advertisements, and malicious links, thus ensuring a more reliable processing outcome for us, and in turn, a more authentic and reliable educational environment on YouTube. The study's innovative use of a graphical user interface (GUI) and a Telegram bot further simplified the process of sentiment analysis for its users, making

it more accessible and user-friendly for stakeholders involved in educational content creation and dissemination.

These findings highlight the importance of sentiment analysis in understanding viewer responses towards educational videos, providing a nuanced view of the audience's engagement levels and content reception. This analysis facilitated the identification of prevailing sentiments and themes, which were visualized through word clouds for a deeper understanding. The performance of various machine learning models, including Multinomial Naive Bayes, Logistic Regression, SVC, Decision Tree, and Random Forest, was thoroughly evaluated, with SVC emerging as the most accurate model for sentiment classification. The data collection through YouTube API integration to comment analysis and model evaluation, demonstrates the effectiveness of the employed tools and algorithms in navigating the complexities of sentiment analysis in an educational context.

Ethical considerations, such as anonymizing comments and adhering to YouTube's terms of service, were meticulously observed, ensuring the integrity of the analysis and interpretation of user-generated content. These ethical practices, alongside the study's methodological soundness, contribute to the validity and reliability of the findings, offering a robust foundation for enhancing educational videos and the viewer learning experience on YouTube. As the study contributes to the broader field of sentiment analysis in user-generated content, it also paves the way for future research to explore advanced NLP techniques, expand the scope of analysis across different video categories, and incorporate additional features for better sentiment analysis. Such continued efforts in sentiment analysis will empower both groups of students and viewers along with content creators, educators, and researchers with deeper insights into audience engagement, driving improvements in content quality and fostering more effective interactions with the online educational content on YouTube.

References

1. Foreman-Brown, G., Fitzpatrick, E., Twyford, K.: Reimagining teacher identity in the post-Covid-19 university: becoming digitally savvy, reflective in practice, collaborative, and relational. Educ. Develop. Psychol. **40**(1), 18–26 (2023)
2. Mazzara, M., et al.: Education After COVID-19. In: Smart and Sustainable Technology for Resilient Cities and Communities, pp. 193–207 (2022)
3. Beautemps, J., Bresges, A.: What comprises a successful educational science YouTube video? A five-thousand user survey on viewing behaviors and self-perceived importance of various variables controlled by content creators. Front. Commun. **5**(600595) (2021). https://doi.org/10.3389/fcomm.2020.600595
4. Bitzenbauer, P., Höfler, S., Veith, J.M.: Exploring the relationship between surface features and explaining quality of YouTube explanatory videos. Int. J. Sci. Math. Educ. **22**, 25–48 (2024)
5. Jelodar, H., et al.: A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on YouTube comments. Multimedia Tools Appl. **80**, 4155–4181 (2021)
6. Cunha, A.A.L., Costa, M.C., Pacheco, M.A.C.: Sentiment analysis of YouTube video comments using deep neural networks. In: Artificial Intelligence and Soft Computing, pp. 561–570 (2019)

7. Jain, J., Dey, D., Kelkar, B., Ahlawat, K.: Analysis of indian news with corona headlines classification. In: Dev, A., Agrawal, S.S., Sharma, A. (eds.) Artificial Intelligence and Speech Technology. AIST 2021. CCIS, vol. 1546. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-95711-7_10
8. Alwash, M., Savarimuthu, B.T.R., Parackal, M.: Mining brand value propositions on Twitter: exploring the link between marketer-generated content and eWOM outcomes. Soc. Netw. Anal. Min. **11**(1), 83 (2021)
9. Zhou, J., Ye, J.-M.: Sentiment analysis in education research: a review of journal publications. Interact. Learn. Environ. **31**(3), 1252–1264 (2023)
10. Le, T.C., Nguyen, Q.V., Tran, M.T.: SmartEyes: social multimedia analysis platform for open data providers. SN Comput. Sci. **3**(1), 33 (2022)
11. Nayak, S., Sharma, Y.K.: A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis. Dec. Anal. J. **8**, 100289 (2023)
12. Harish, B.S., Rangan, R.R.K.A.: Comprehensive survey on Indian regional language processing. SN Appl. Sci. **2**(7), 1204 (2020)
13. Mamun, M.M.R., Sharif, O., Hoque, M.M.: Classification of textual sentiment using ensemble technique. SN Comput. Sci. **3**(1), 49 (2022)
14. Sahu, M.K., Selot, S.: Comparative analysis of various supervised machine learning techniques used for sentiment analysis on tourism reviews. In: Proceedings of International Conference on Recent Trends in Computing: ICRTC 2021, pp. 19–49. Springer Singapore (2022)
15. Dogra, V., Singh, A., Verma, S., Kavita, N., Jhanjhi, Z., Talib, M.N.: Analyzing DistilBERT for sentiment classification of banking financial news. In: Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021, pp. 501–510. Springer Singapore (2021)
16. Jim, J.J.R., Talukder, M.A.R., Malakar, P., Kabir, M.M., Nur, K., Mridha, M.F.: Recent advancements and challenges of NLP-based sentiment analysis: a state-of-the-art review. Nat. Lang. Process. J. 100059 (2024)
17. Hidayat, T.H.J., Ruldeviyani, Aditama, Y., Madya, G.R., Nugraha, A.W., Adisaputra, M.W.: Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. Procedia Comput. Sci. **197**, 660–667 (2022)
18. M. Amini Motlagh, H. Shahhoseini, and N. Fatehi, "A reliable sentiment analysis for classification of tweets in social networks," *Social Network Analysis and Mining*, vol. 13, no. 1, p. 7, 2022
19. Madasu, A., Elango, S.: Efficient feature selection techniques for sentiment analysis. Multimedia Tools Appl. **79**(9), 6313–6335 (2020)
20. Khaund, T., Hussain, M.N., Shaik, M., Agarwal, N.: Telegram: Data collection, opportunities and challenges. In: Lossio-Ventura, J.A., Valverde-Rebaza, J.C., Díaz, E., Alatrista-Salas, H. (eds.) Information Management and Big Data. SIMBig 2020. CCIS, vol. 1410. Springer, Cham (2010). https://doi.org/10.1007/978-3-030-76228-5_3

Enhancing Crop Yield Through Convolutional Neural Network (CNN) Powered Plant Disease Detection

Kalyani Satone(✉) and Pranjali Ulhe

SVPCET, Nagpur, India
{ksatone,pranjalideshmukh}@stvincentngp.edu.in

Abstract. As the global population continues to grow, the demand for agricultural production is on the rise. Plant diseases pose significant challenges to the agriculture industry and its management. Historically, humans have relied on visual recognition to identify plant diseases, a process often marred by subjectivity and time constraints. To streamline disease recognition, machine learning techniques leveraging images of plant leaves have emerged. Timely detection is crucial as these diseases can profoundly impact a plant's development. Crop loss due to pathogens like bacteria, viruses, and fungi has plagued agriculture for centuries on a global scale. The widespread availability of smartphones and recent advancements in deep learning-based computer vision have paved the way for smartphone-based disease detection. We have employed a dataset consisting of diverse images of both healthy and diseased plant leaves captured under controlled conditions to train a deep convolutional neural network. These breakthroughs have now opened the door to global-scale mobile-phone-based plant disease diagnosis, particularly with the advent of larger publicly available image datasets and advancements in Artificial Intelligence (AI) and Deep Learning (DL). This research demonstrates the significant potential for improved accuracy in this field, as AI and DL, a subset of Machine Learning (ML), continue to advance. Various ML models have been applied to the task of identifying and classifying plant diseases, and this work provides a comprehensive exploration of DL models tailored to represent a wide spectrum of plant diseases.

Keywords: Artificial intelligence · deep learning · convolutional neural networks (CNN) · plant disease detection · and machine learning

1 Introduction

Advanced crop disease detection and prevention not only improves productivity and ensures agricultural sustainability, but also minimizes disease-related damage to crops during growth, harvest, and post-harvest treatment. Essential to curb it. Some of the direct detection techniques include

- Gas chromatography-mass spectrometry (GC-MS)
- immunofluorescence (IF)

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 170–180, 2026.
https://doi.org/10.1007/978-3-031-98138-8_14

- hybridization using fluorescence in situ (FISH)
- enzyme-linked immunosorbent assay (ELISA)
- flow cytometry (FCM)
- polymerase chain reaction (PCR)

Examples of indirect methods include hyperspectral approaches, fluorescence imaging, and thermography. Farmers typically cannot use these methods as certain knowledge is required to use them (Fig. 1).

Fig. 1. Diseased Crop in Natural Setting

Global food security is at risk from plant diseases, but small - scale farmers whose economy is dependent on growing healthy crops can also suffer catastrophic consequences. More than 80% of agricultural production in poor countries is produced by small farmers, with yield losses of more than 50% due to pests and diseases often claimed (UNEP 2013). (Harvey et al. 2014). Small farmers, a group particularly vulnerable to disruption of food supplies by pathogens, make up 50% of the world's hunger-stricken households (Fig. 2).

To stop agricultural loss caused by diseases, numerous programmes have been developed. Integrated pest management (IPM) tactics have replaced traditional methods of applying insecticides widely over the past ten years (Ehler 2006). Whatever the method, the first step in effective illness management is accurate disease identification when it first manifests. Historically, institutions like local plant clinics or agricultural extension agencies have assisted in disease identification. These programmes have recently benefited from the accessibility of online tools for illness identification, which take advantage of the rising global internet usage. More recently, the use of technology for mobile phones has expanded thanks to historically unparalleled global adoption (ITU 2015).

Hence, to ease farmer's work, we propose a Deep Learning (DL) Model-based approach that is simple to use and easily accessible for diagnosing plant leaf diseases.

Fig. 2. Experimentally Diseased Plants in a Controlled Environment

2 Literature Review

Author [1] used a neural network classifier for classification. They obtained an accuracy of 97.30% . The main drawback is that it is used only on limited crops. The dataset consists of labeled images of healthy pomegranate leaves and leaves affected by different diseases. Paper uses the concept of neural network architecture used for disease classification. It included the number of layers, neurons in each layer, activation functions, and the training algorithm used for the neural network. The paper described the process of feature extraction from the pomegranate leaf images, which are fed into the neural network for classification. This step is crucial for converting the raw image data into relevant input features for the neural network. The paper would likely evaluate the performance of the neural network model in classifying different pomegranate diseases. Metrics such as accuracy, precision, recall, and F1-score might be used to assess the model's effectiveness.

R. V. Kshirsagar [2] using snake segmentation typically refers to active contour models, also known as snakes, which are used for image segmentation tasks. These models use energy minimization techniques to find the boundaries of objects in an image. They have been applied in various fields, including medical imaging, computer vision, and object recognition. The average classification is 85.52%.

Siddharth et al. [3] has been introduced a model to identify and classifying the diseases in plant leaf images. The model basically depends on Radial Bias Function Neural Network. Authors provided the algorithm to search a common attribute by grouping the seed points to find the features. The specificity of the proposed model for segmentation is 0.558 and for classification upon Vpc and Vpe is 0.8621 and 0.1118 respectively.

Aydin et al. [4] has studied different learning models to evaluate the performance. Authors has mentioned four models as CNN, LDA, AlexNet and VGG16 on available datasets. They had taken the five fold cross-validation procedure and considered the size of an image as 100. The proposed model obtained an accuracy score of 96.93% whereas the pre-trained VGG16 model outperformed with an accuracy of 99.80%.

Author [4] tried a technology, used an artificial neural network (ANN) as a classifier. This is useful for checking the severity of diseased leaves. The research paper is likely to mention the dataset used for training and testing the disease detection and grading system.

The dataset consist of labeled images of healthy leaves and leaves affected by various diseases, along with corresponding severity or grading labels. The paper described the computer vision techniques used to extract features from leaf images. This could include image preprocessing, feature extraction algorithms, and methods for representing leaf disease patterns effectively. The paper might discuss the design of the fuzzy system, membership functions, rules, and defuzzification methods used to convert image-based features into disease severity grades.

The core of the paper would be the explanation of the overall disease detection and grading system, which combines computer vision technology and fuzzy logic. It describes the integration of the two approaches and how they work together to achieve accurate disease detection and grading. This paper presented the results obtained from the disease detection and grading system and discussed the system's performance compared to other methods or baselines. The authors compared the strengths and limitations of their approach and discussed potential areas for improvement.

Saleem, Potgieter, and Arif [7] detected the concept of using deep learning algorithms for automating the process and improving accuracy in plant disease identification.

They have included Convolutional Neural Networks (CNNs) or more advanced architectures like ResNet, Inception, or DenseNet. They have maintained the Transfer Learning and Pre-trained Models .When dealing with limited labeled data. While deep learning has shown promise in plant disease detection, several challenges persist, including the need for large-scale labeled datasets, model interpretability, and dealing with multi-class imbalanced datasets.

By summarizing the key findings from Saleem et al.'s study and other related research. It emphasized the potential of deep learning as a valuable tool for accurate and efficient plant disease detection and classification, while acknowledging the need for further research to optimize and adapt these models for real-world agricultural settings

Author [11] suggested key findings and highlighting the significance of using CNNs for agricultural remote sensing image classification. It also emphasizes the importance of continuous research and development in this area to unlock the full potential of CNNs in improving agricultural practices and sustainability.

M. Sardogan [12] focusing on plant leaf disease detection using deep learning and other machine learning techniques with different classification algorithms is presented. The strengths and weaknesses of various approaches are highlighted.

Author [13] summarizes the progress made in utilizing deep learning CNN models for plant disease detection using image segmentation. It highlights the strengths and weaknesses of existing models and identifies potential avenues for future research to enhance the accuracy and robustness of these models.

3 Methodology

We are Considering 61,486 photographs of plant leaves with a total of 39 class designations. We try to predict plant-disease pairs for each class label based only on plant leaf images. Each class designation represents a plant-disease pair. Each of the research techniques involves downscaling the images to 256 $\times$ 256 pixels, model optimization, and prediction on those images, followed by taking the augmented data sets.

Large datasets are typically used by DL algorithms to solve issues like overfitting. When introducing algorithms for more general practical use, this frequently presents a barrier. The process of collecting data can be time-consuming, and the labeling activities may call for the assistance of subject-matter specialists, to overcome this problem augmentation techniques are used. These techniques are a popular strategy to increase the size of the current datasets without losing its uniqueness. Deep learning models demonstrate greater accuracy with large datasets.

The size of the filters and outputs changes as the data moves through the network (becoming smaller as the data passes through networks), enabling the training and identification of identical features with varied scaling. Due to the fact that convolution filters were applied to the complete collection of images used as training data, the CNNs are invariant to changes in properties like rotation and translation. Additionally, feature distortions are successfully corrected by the pooling layers in the CNNs (Figs. 3, 4 and 5).

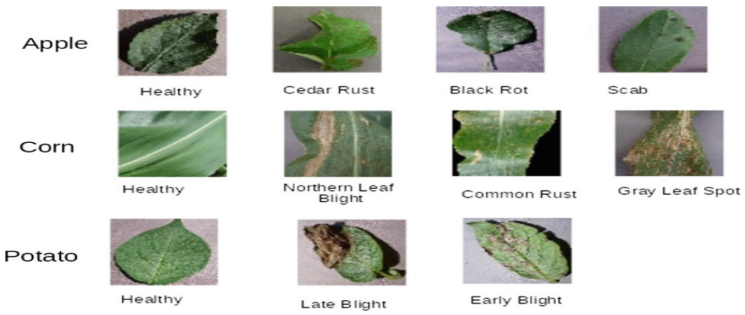


Fig. 3. Healthy and non-healthy leaf from Apple, Corn, or Potato Plant

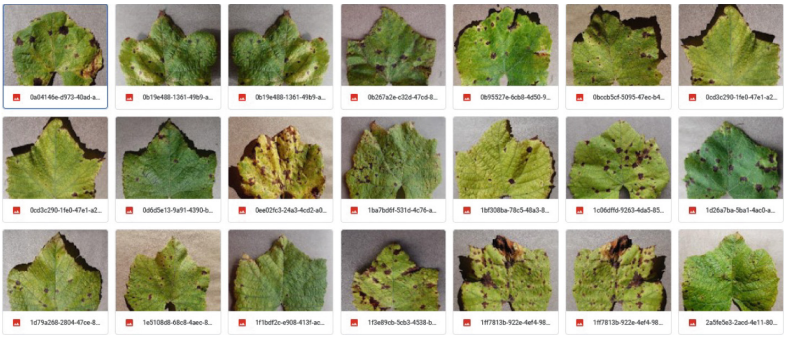


Fig. 4. Image Dataset for Digital Image Analysis

Following the dataset upload, the model is trained and validated. We used libraries like Tensorflow, Pathlib, Keras, and Torch to create the model. The model was created using the ReLU (Rectified Linear Unit) Activation Function along with Adam optimizer.

index	disease_name	description
0	Apple : Scab	Apple scab is the most common disease of apple and crabapple trees in Minnesota. Scab is caused by a fungus that infects both leaves and fruit. Scabby fruit are
1	Apple : Black Rot	Leaf symptoms first occur early in the spring when the leaves are unfolding. They appear as small, purple specks on the upper surface of the leaves that enlarge i
2	Apple : Cedar rust	Cedar apple rust (Gymnosporangium juniperi-virginianae) is a fungal disease that requires juniper plants to complete its complicated two year life-cycle. Spores c
3	Apple : Healthy	As with most fruit, apples produce best when grown in full sun, which means six or more hours of direct summer Sun daily.The best exposure for apples is a nortl
4	Background Without Leaves	There is no leaf in the given image
5	Blueberry : Healthy	Blueberries Are Low in Calories But High in Nutrients. ... Blueberries are the King of Antioxidant Foods. ... Blueberries Reduce DNA Damage, Which May Help Prot
6	Cherry : Powdery Mildew	Initial symptoms, often occurring 7 to 10 days after the onset of the first irrigation, are light roughly-circular, powdery looking patches on young, susceptible leav
7	Cherry : Healthy	There is no difference in care between sour and sweet cherries. Apply mulch to retain moisture. Drape netting over trees to protect the fruit from birds. Water rou
8	Corn : Cercospora Leaf Spot | Gray Leaf Spot	Gray leaf spot on corn, caused by the fungus Cercospora zeae-maydis, is a peren- nial and economically damaging disease in the United States. Since the mid-19
9	Corn : Common Rust	Although a few rust pustules can always be found in corn fields throughout the growing season, symptoms generally do not appear until after tasseling. These ca
10	Corn : Northern Leaf Blight	Northern corn leaf blight (NCLB) is caused by the fungus Setosphaeria turcica. Symptoms usually appear first on the lower leaves. Leaf lesions are long (1 to 6 inc
11	Corn : Healthy	Corn plants prefer daytime temperatures of 75 to 80 degrees F and 65 to 70 degrees F during the night. The soil should be kept consistently moist, but not soggy
12	Grape : Black Rot	Grape black rot is a fungal disease caused by an ascomycetous fungus, Guignardia bidwellii, that attacks grape vines during hot and humid weather. "Grape black
13	Grape : Esca | Black Measles	Grapevine measles, also called esca, black measles or Spanish measles, has long plagued grape growers with its cryptic expression of symptoms and, for a long ti
14	Grape : Leaf Blight | Isariopsis Leaf Spot	The fungus is an obligate pathogen which can attack all green parts of the vine. Symptoms of this disease are frequently confused with those of powdery mildew
15	Grape : Healthy	Apply water only to the root zone. Avoid getting grape foliage wet as this can encourage many grape diseases. Reduce watering young vines in the fall to encour
16	Orange : Haunglongbing | Citrus Greening	Citrus greening disease is a disease of citrus caused by a vector-transmitted pathogen. HLB is distinguished by the common symptoms of yellowing of the veins
17	Peach : Bacterial Spot	Bacterial spot is an important disease of peaches, nectarines, apricots, and plums caused by Xanthomonas campestris pv. pruni. Symptoms of this disease include

Fig. 5. Comprehensive Plant Disease Encyclopedia Dataset

Different pre-processing approaches are taken into consideration to eliminate noise in images or other object removal. To obtain the desired image region, an image of a leaf is cropped, or clipped as per the requirement (Fig. 6).

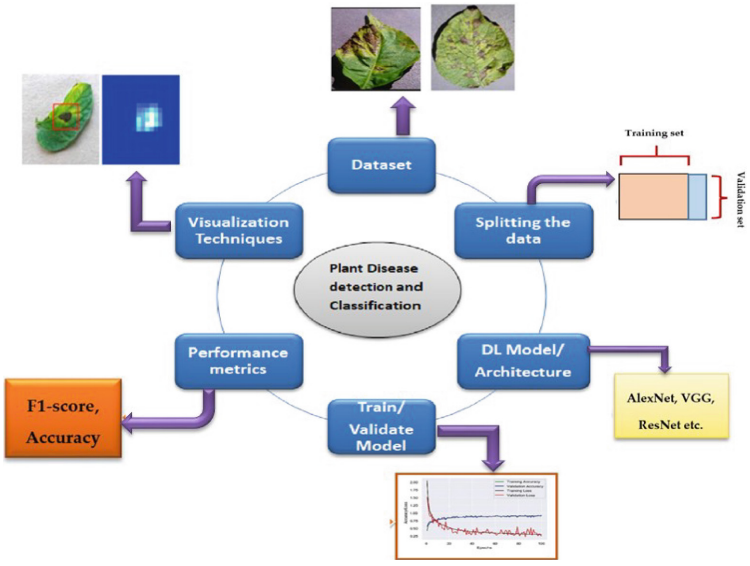


Fig. 6. Different phases representing methodology for plant disease detection

The smoothing filter is used to smear an image. For the purpose of boosting contrast, images are enhanced.

Adam combines the best elements of the RMSProp and AdaGrad algorithms to create an optimization strategy that can handle sparse gradients in noisy problems. All segmentation, annotation, and labeling techniques required selecting only lesions of interest, applying random distortion, blurring, brightness, and contrast changes, and assigning individual labels to each. CNN Random modifications (stretching, rotation, brightness, contrast blurring) were made to each segmented image before testing with the model.

Result: Some plants had self classification accuracy more than 92%, while in other cases accuracy was too low due to insufficient data. For example, Spider Mite was discovered as the class in about 3% of photos with Target Spot. This is due to a lot of Spider Mite photos also having Target Spot symptoms, which confused the model during training. The similarity of the photos used for both training and validation in the Plant Village data can be partly credited for this great accuracy. The model also has a serious flaw, when tested on the same shot with only half a leaf cropped, or another image from the web, the average self-classification accuracy drops to just 52.3%. This is consistent with other work that trained the algorithm on photos from the Plant Village collection and compared the results to independent images (Figs. 7, 8, 9, 10 and Tables 1, 2).

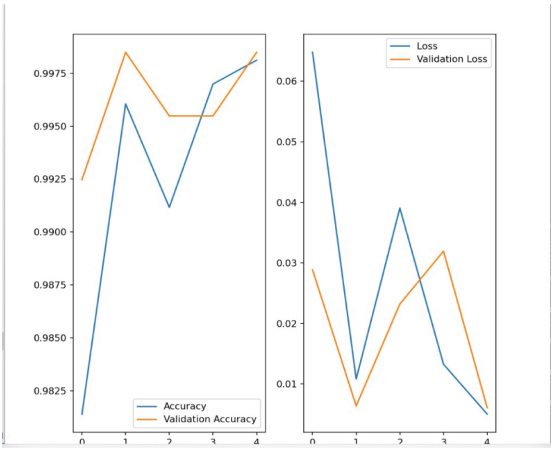


Fig. 7 Orange plant disease detection accuracy

Conclusion: The proposed system detects diseases to a certain extent. The performance of most deep learning models for autonomous disease detection degrades when applied to real, unexplored images. We trained a convolutional neural network (CNN) model using segmented and annotated images rather than whole images. The performance of the same CNN model on independent data improves when segmented images are used for training as opposed to whole images.

However, there are still a number of issues that need to be resolved in further work. The accuracy of the model is first noticeably diminished when tested on a batch of

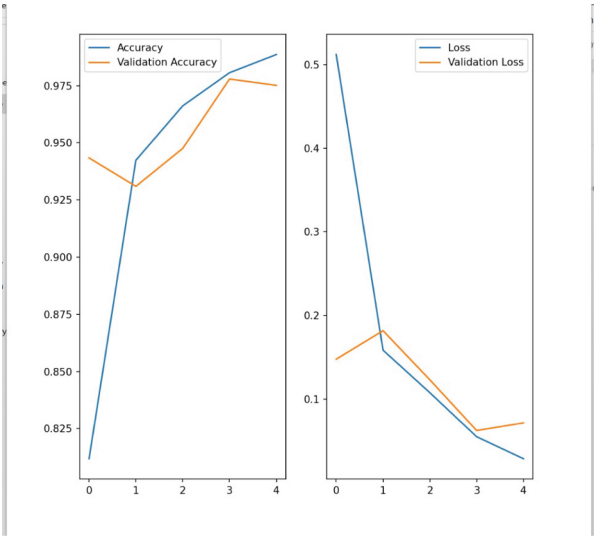


Fig. 8. Apple plant disease detection accuracy

Table 1. Analysis of Accuracy and Validation Accuracy

points	Accuracy	Validation Accuracy
0	Null	0.8
1	Null	0.9
2	10%	0.9
3	20%	0.9
4	30%	0.9

photographs acquired in conditions that differ from those used for training. This accuracy is much better than that based on a randomly selected subset of 38 classes, but more diverse training data is needed to improve the accuracy. Efforts are being made to collect more (and more variable) data in light of recent results showing significant improvements in accuracy.

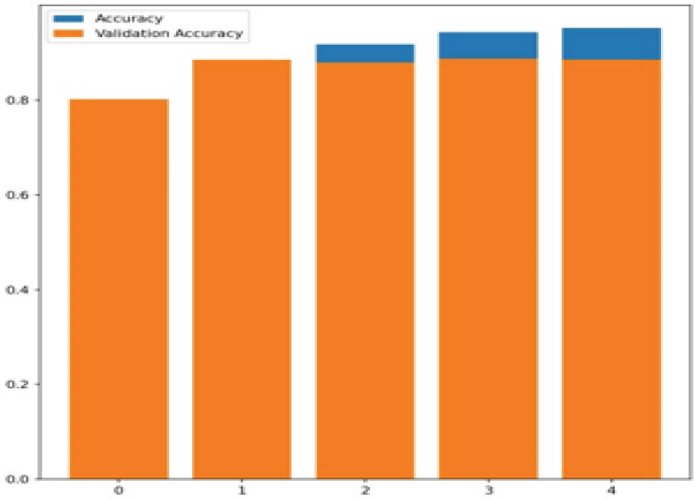


Fig. 9. Graphs representing accuracy and validation accuracy

Table 2. Analysis of Loss and Validation Loss

points	Loss	Validation Loss
0	30%	.625
1	10%	.385
2	Null	0.4
3	Null	.385

Because growers are expected to be knowledgeable about the crops they produce, the use of 38 classes, including combined plant species and pathologies, has made the problem more complicated than necessary. Due to the extraordinarily high accuracy of the Plant Village dataset, restricting the classification task for each disease does not have a large impact. We can see that the accuracy is significantly improved on the real dataset. Overall, the provided method performs quite well with a wide range of plant species and diseases, and it is anticipated that with more training data, it would perform even better.

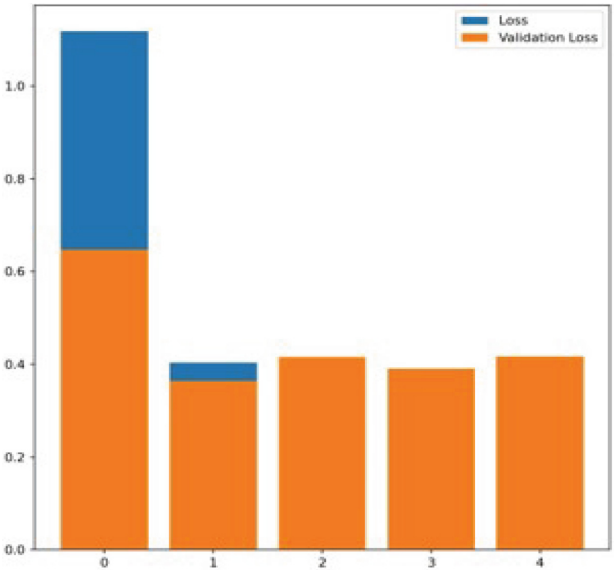


Fig. 10. Graphs representing loss and validation loss

References

1. Sannakki, S.S., Rajpurohit, V.S.: Classification of pomegranate diseases based on back propagation neural network. Int. Res. J. Eng. Technol. **2**(2) (2015)
2. Rothe, P.R., Kshirsagar, R.V.: Cotton leaf disease identification usingpatternrecognition techniques. In: 2015 International Conference on Pervasive Computing (ICPC), pp. 1–6 (2015). https://doi.org/10.1109/PERVASIVE.2015.7086983
3. Rastogi, A., Arora, R., Sharma, S.: Leaf disease detection and grading using computer vision technology & fuzzy logic. In: 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 500–505 (2015). https://doi.org/10.1109/SPIN.2015.7095350
4. Fang, Y., Ramasamy, R.P.: Current and prospective methods for plant disease detection. Biosensors **5**(3), 537–561 (2015)
5. Jeong, S., Jeong, S., Bong, J.: Detection of tomato leaf miner using deep neural network. Sensors **22**, 9959 (2022). https://doi.org/10.3390/s22249959
6. Saleem, M.H., Potgieter, J., Arif, K.M.: Plant disease detection and classification by deep learning. Plants **8**(11), 468 (2019)
7. Mishra, R.K., Urolagin, S., Gaur, J.A.A.J.P.: Deep hybrid learning for facial expression binary classifications and predictions. Image and Vision Computing (2022)
8. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016)
9. Kruse, O.M.O., Prats-Montalbán, J.M., Indahl, U.G., Kvaal, K., Ferrer, A., Futsaether, C.M.: Pixel classification methods for identifying and quantifying leaf surface injury from digital images. Comput. Electron. Agric. **108**, 155–165 (2014)
10. Hughes, D., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 (2015)

11. Yao, C., Zhang, Y., Zhang, Y., Liu, H.: Application of convolutional neural network in classification of high resolution agricultural remote sensing images. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **XLII-2/W7**, 989–992 (2017)
12. Sardogan, M., Tuncer, A., Ozen, Y.: Plant leaf disease detection and classification based on CNN with LVQ algorithm. In: 2018 3rd International Conference on Computer Science and Engineering (UBMK), pp. 382–385 (2018). https://doi.org/10.1109/UBMK.2018.8566635
13. Sharma, P., Berwal, Y.P.S., Ghai, W.: Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. Inform. Process. Agric. **7**(4) (2020)

A Systematic Review on Anomaly Detection Techniques for Fog Computing Devices

Gourav Mondal[1,2(✉)] and Rajesh Kumar Dhanaraj[1]

[1] Symbiosis International (Deemed) University, Pune, India
gourav.ju@gmail.com, Sangeraje@gmail.com
[2] Netaji Subhash Engineering College, Kolkata, India

Abstract. Background: The article provides an overview of existing techniques, their advantages and disadvantages, and their applicability in fog computing. Anomaly detection techniques in fog computing devices are crucial for ensuring the reliable and safe operation of fog computing environments.

Problem: The scattered and heterogeneous nature of fog computing, on the other hand, creates a number of concerns, including possible security risks and operational uncertainty. Furthermore, the paper outlines numerous outstanding issues and research objectives, such as enhancing the approaches' accuracy, scalability, and resilience, creating techniques that can effectively identify multiple types of anomalies, and developing strategies that can be used in real-world scenarios.

Scope: Anomaly detection techniques play a pivotal role in mitigating these challenges by identifying unusual behaviors that may indicate security breaches, faults, or deviations from normal operational patterns. Furthermore, the review identifies several open challenges and research directions, such as improving the accuracy, scalability, and robustness of the techniques, developing techniques that can efficiently detect multiple types of anomalies, and developing techniques that can be deployed in large-scale fog computing networks. This review gives useful insights into existing anomaly detection approaches. To handle IoT networks, Fog computing systems require unique SLAs, bandwidth-aware design, and scalability. Fog-resource monitoring, green computing, and Federated Reinforcement Learning (FRL) can improve energy consumption and reliability.

Aim: The aim of anomaly detection for fog computing devices is to enhance the overall performance, privacy, security, and stability of fog computing environments by promptly identifying and addressing anomalies.

Keywords: Anomaly detection · IOT · Fog Computing · Genetic algorithms · Fuzzy Logic · Bayseian Networks

1 Introduction

An overview of the history of anomaly detection methods for fog computing devices is given below: Various Anomalies: Point anomalies (single occurrences that significantly deviate from the norm), contextual anomalies (occurrences that are normal in isolation

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 181–196, 2026.
https://doi.org/10.1007/978-3-031-98138-8_15

but abnormal in the context in which they occur), and collective anomalies (collectives of occurrences that are normal individually but abnormal when taken together) are the three main categories into which anomalies can be divided. All of these kinds of abnormalities must be addressed by fog computing systems.

Data Sources and Characteristics: Fog computing devices are network-connected devices that generate massive volumes of data from numerous sources such as IoT sensors, wearables, surveillance cameras, and more. Data properties, such as organized, semi-structured, and unstructured data, time-series data, and sensor readings, can vary greatly. These various data types must be handled by anomaly detection algorithms.

Processing in real time: The capacity of fog computing to handle real-time applications is one of its primary advantages. Fog device anomaly detection systems should be efficient enough to examine data streams in real- time, giving quick alerts and answers to suspected anomalies.

Machine Learning and AI-Based Approaches: Machine learning techniques, particularly supervised and unsupervised learning, are widely used in fog computing for anomaly identification. To distinguish between normal and abnormal cases, supervised learning systems require labeled training data. Unsupervised learning algorithms, on the other hand, can detect abnormalities in the absence of labeled data by learning normal behavior and identifying deviations.

Fog computing systems are naturally scattered, with numerous units spread across multiple areas. To detect global abnormalities and trends, this dispersion necessitates anomaly detection techniques that can work autonomously on each device while also collaborating with other devices and central management systems.

Fog devices frequently have limited computational power, memory, and battery resources. Techniques for detecting anomalies must be resource-efficient and intended to function successfully in resource-constrained contexts.

A combination of rule-based approaches and machine learning technologies may be utilized to detect anomalies more accurately. Machine learning algorithms can capture complicated patterns and anomalies that are difficult to represent directly, whereas rule-based systems can detect basic abnormalities based on predetermined thresholds.

As the fog environment changes and new data is collected, anomaly detection algorithms must adapt and learn from new patterns and behaviors. Continuous learning strategies ensure that the models stay relevant and effective over time.

For IoT applications, fog models scatter streams over networks. The hierarchical plane stack of Fog handles user-edge data. The IoT platform takes advantage of cloud computing because it is inexpensive and time- consuming. As a result, fog computing is advised for time-critical processes [1].

Smart parking, health, traffic control, energy distribution, supplies and other resources, and supply chain management are examples of IoT applications. These programs all monitor and react to sensor conditions. Some observations are crucial, while others may astound you. These surprising discoveries could be the consequence of climate change, human intervention, system failure, or pure luck. They were misfits. The limitations of edge sensors increase the likelihood of anomalies [2].

The ability to detect errors early, fix irregularities, and adapt to changing situations are critical to preserving method validity. The detection of cloud anomalies has been

successful. Transferring submitted data to a distant server to search for outliers requires a substantial amount of bandwidth when billions of sensors are used. Peripheral systems cannot do so. Many devices, however, do not have cloud connectivity. The detection of irregularities diminishes visibility in fog or the detection time of the edge layer [3].

Figure 1 illustrates the fog network's heart. Fog provides cloud computing to customers by addressing transportation, latency, and network constraints. This is known as "IoT + data and business-designed insights derived from data transmitted among these smart gadgets." "The fog nodes are more responsive than the cloud" and may be expanded to meet hosting and security needs. Its cloud-like tiered foundation might support a wide range of applications. Fog computing, as opposed to sending data to the cloud or another centralized point, conserves network resources and reaction time.

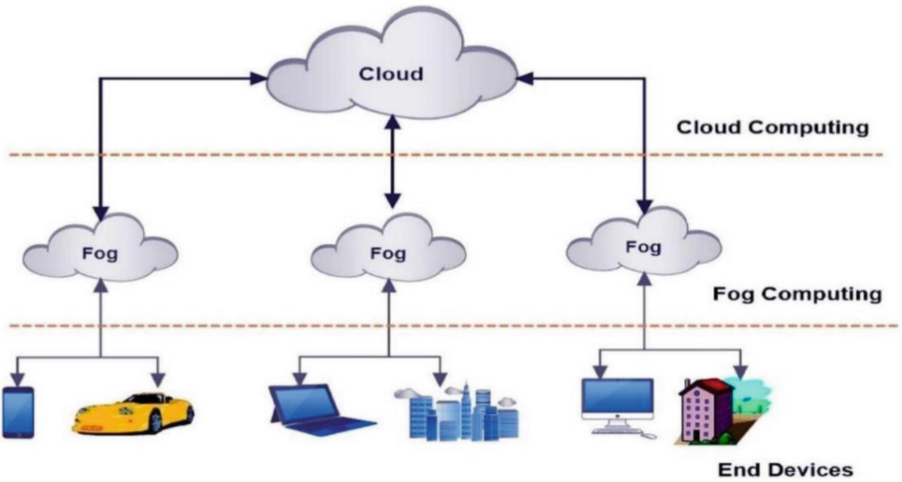


Fig. 1. Modified form of Fog Structure [5]

The remainder of the paper is structured as follows. Section 2 investigates ways for detecting background anomalies. Section 3 tabulates and describes the literature review in depth. Section 4 discusses fog computing and its architecture. Section 5 discusses the various methods of anomaly detection. Section 6 analyzes several approaches of anomaly detection while discussing fog computing. Section 7 focuses on challenges and possible solutions.

1.1 The History of Anomaly Detection Techniques

Outlier detection is the search for anomalies or things that are significantly out of the norm. Any anomaly detection relies on two essential assumptions, which data scientists frequently apply to unlabeled data in a process known as unsupervised anomaly detection [6].

- Data anomalies are infrequent.
- Data anomalies exhibit features that are unique from regular occurrences.

Anomaly data concerns include hacking, bank fraud, broken equipment, structural defects/infrastructure failures, and typographical mistakes. It is critical in business to discern genuine irregularities from false positives and data noise.

Finding and recording unusual occurrences, things, or observations that deviate from the norm is what "anomaly detection" entails. Data anomalies can take numerous forms, including standard deviation (sd), outlier, turbulence, and so on**.** Network attacks and other types of network intrusion and abuse are uncommon but not unheard of. Unique activity bursts stand out despite their statistical obscurity.

Unsupervised outlier identification algorithms may miss such a sudden increase in activity. Cluster analysis, on the other hand, frequently discovers tiny micro-clusters. The primary techniques are unsupervised, semi- supervised, and supervised anomaly detection. The optimum anomaly detection strategy is determined by dataset labeling.[8].To classify supervised anomaly detection algorithms, a data set labeled "normal" and "abnormal" is required. This method necessitates classifier training. Unlike typical pattern recognition, outlier detection has a significant class imbalance. Because anomaly detection is unequal, not all statistical classification approaches work.[9].

Semi-supervised anomaly detection systems replicate normal behavior using labeled training data. They put the model's capacity to replicate to the test. Unsupervised anomaly detection may uncover abnormalities in unlabeled test sets. These strategies rely only on data. As is customary, the majority of data points will be normal. Anomaly detection will find instances that differ from the data set.[10].

2 Literature Review

The information on the literature review is provided below in Table 1.

Table 1. Various writers' synopses of their reviews of the relevant literature

Author	Title	Methodology	Outcomes	Future work
[11]	Smart Meter Data DistributedFog Computing Architecture for Real-Time Anomaly Detection	o implement anomaly detection models based on machine learning for residential Smart Homes, a hierarchically Distributed Fog Computing architecture is being developed	This architecture is more effective than Cloud computing because of big data, global awareness, and low latency	We may test our real-time anomaly detection system on several Edge layer devices with diverse setups

(*continued*)

Table 1. (*continued*)

Author	Title	Methodology	Outcomes	Future work
[12]	Anomaly detection in a fogcomputing environmentusing genetic algorithms and Nave Bayes	Using the Security Laboratory Knowledge Discovery Dataset, a better model was constructed that can predict outcomes more accurately by reducing extraneous variables to reduce time complexity and applying the Nave Bayes for Anomaly DetectionModel (GANBADM) in a Fog Environment (NSL-KDD)	The created method has99.73% accuracy and 0.6% false positives	Use real cloud data to test this strategy
[13]	Anomalydetection framework to prevent DDoS attacks in fog-empoweredIoT networks	presented a fog layer-based IoT anomaly detection framework	Theproposed method greatly lowers network bandwidth use when comparedto centralized cloud solutions	Prototyping and testing will be done withourfog computing technology
[14]	An IoT Anomaly Mitigation Framework Based on Fog Computing	A fog computing architecture for 5G-enabled smart city autonomous management and orchestration was proposed	The model is more robust and collects more network traffic statistics	Including network metrics in the anomaly detection model, such as energy consumption, source IPs, etc
[15]	Fog Computing: Managing and Orchestrating Smart City Applications in 5G Networks	offer a hybrid IoT anomaly mitigation solution based on fog computing for faster and more accurate detection.	Signature-based modules detect attacks more quickly than anomaly-based modulas. The anomaly-based module is also effective in detecting attacks	Attempt to investigate more botnet attack features in order to develop more framework signatures.

3 Devices and Architecture for Fog Computing

In this section, we'll take a high-level look at what fog computing is and how it's built.

3.1 Fog Computing Definition

The use of fog computing lacks the complete set of functions of typical cloud computing, such as the capacity to compute, store, and network services across multiple end devices. It's a feasible solution for latency- sensitive Internet of Things applications [16].

Despite the fact that fog computing was invented, numerous scholars and organizations have used different terms to describe it. [18] has defined the term "fog computing" liberally. "Fog Computing is a geographically distributed computing architecture that provides elastic compute, storage, and communication to a dense concentration of nearby customers in isolated settings." "Fog Computing is a geographically distributed computing architecture with a resource pool made up of one or more ubiquitously connected heterogeneous devices (including edge devices) at the network's edge that is not exclusively seamlessly backed by Cloud services." While Vaquero and others define fog computing as "a scenario in which a large number of heterogeneous (wireless and sometimes autonomous) ubiquitous and decentralized devices communicate and potentially cooperate among themselves and with the network to perform storage and processing tasks without the intervention of third parties," I prefer to call it a "situation" where devices can communicate. These responsibilities might include network maintenance or the deployment of experimental services and applications. Users that lease equipment to host these services make money. According to the OpenFog Consortium, this is "a system-level horizontal architecture that distributes resources and services of computing, storage, control, and networking anywhere along the continuum from the Cloud to the Things" [20].

3.2 Architecture of Fog Computing

"Fog computing" shifts data center responsibilities to the network's edge. As a result, the fog's limited processing power is divided between end-user devices and faraway cloud servers—tools for storing and connecting data. Low-latency IoT applications are prioritized by fog computing [21].

The six layers depicted in Fig. 2 are the fundamental building parts of fog computing's infrastructure shown by [22] and [23]: storage (both actual and virtual), processing, transmission, protection, and temporary storing and transferring.

Sensor networks are located in the physical and virtualized layers. Upkeep is determined by node type and service. Regional sensors provide environmental data to gateways for analysis and modification. Monitors the behavior of sensors, fog, and network

nodes. This layer keeps track of the current and future duties of network nodes. Performance and uptime of infrastructure services and applications are monitored. Because fog computing includes devices with varying power usage, tracking fog node energy use may improve energy management [24].

It deals with data. Data may be filtered and trimmed to provide useful information. After cloud uploading, the intermediate storage layer keeps preprocessed data momentarily and deletes intermediate storage devices. Data is protected by encryption, and data integrity prevents tampering. After preprocessing, data is delivered to the cloud for service development. Data uploading to the cloud conserves resources. Before delivering data to the cloud, gateway devices process it. The term "smart gateway" best represents this arrangement. Cloud-based smart gateways send sensor network and IoT data to the cloud. New services are fueled by cloud data. Fog computing communication methods [54] must be efficient, lightweight, and adaptive because to restricted resources. Before selecting a communication mechanism, consider the fog's use case [25].

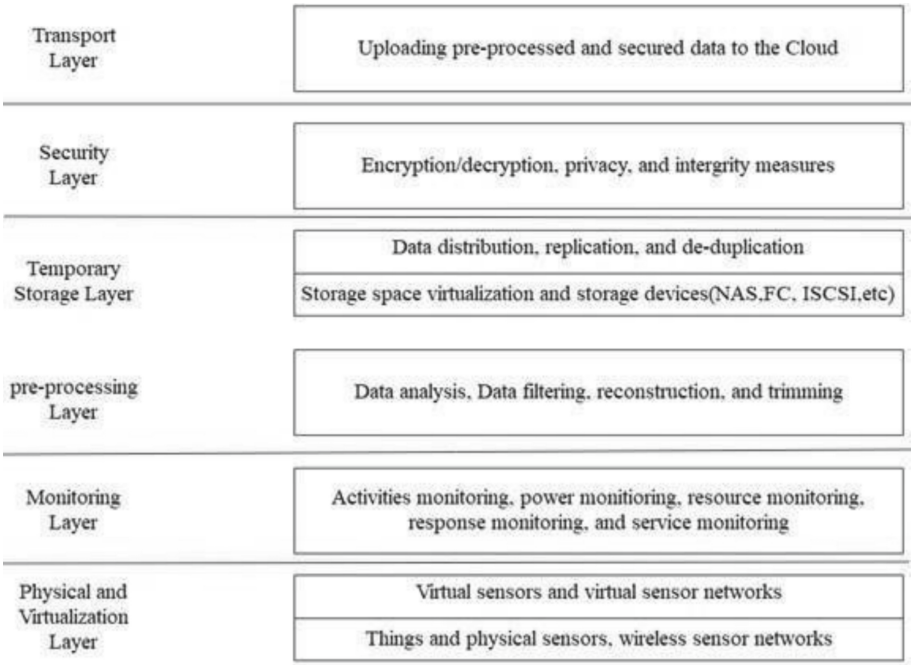


Fig. 2. Fog computing's layered design

4 Fog Computing Device Anomaly Detection Techniques

When looking for data, the user will constantly encounter a lot of noise that might be misinterpreted as abnormalities. Because bad actors vary their tactics, the barrier between normal and aberrant conduct is becoming increasingly blurred. Many data trends are time and season dependent, making it difficult to identify anomalies. Dissecting various trends

over time necessitates more complex strategies for distinguishing true seasonal changes from background noise or outliers.[26].

Many anomaly detection techniques are used in these explanations. One may be more appropriate for a given individual or data collection. To analyze a test case, a generative approach first develops a model using training data samples. Discriminative approaches distinguish between "normal" and "abnormal" data. Both types of data are used in discriminative system training [27].

4.1 Machine Learning-Based Anomaly Detection Techniques

As seen in Fig. 3, statistical, cognitive, and machine learning approaches are used to detect abnormalities. Today, machine learning is used to tackle a wide range of real-world problems. Patterns are classified using machine learning using explicit or implicit models. This topic includes outlier identification, fuzzy logic, Bayesian networks, and genetic algorithms [28].

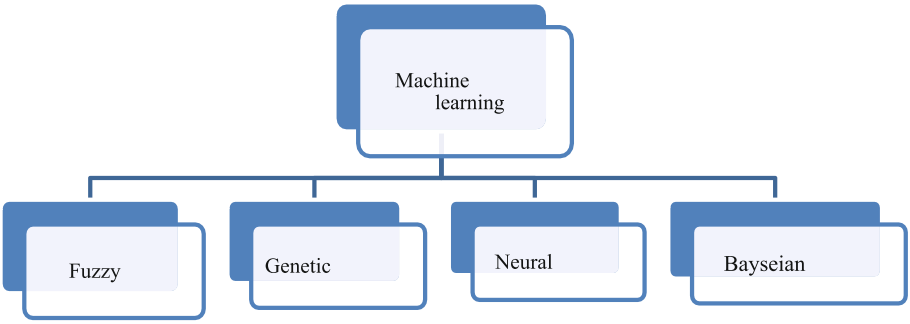


Fig. 3. Machine learning-based techniques categories

4.2 Fuzzy Logic

To produce fuzzy logic, fuzzy set theory approximates classical predicate logic. Because the characteristics to be examined are fuzzy variables, anomaly detection employs fuzzy approaches. Fuzzy logic is effective against port scans and probes, although it is resource intensive [29].

According to Wu and Banzhaf, fuzzy logic identifies network anomalies for two reasons [28]. The initial anomaly detection problems necessitate the collection and statistical determination of several quantitative variables, which may result in a detection error. Second, models that generate typical network activity must define the difference between normal and aberrant activity. This time is not well-defined, and minor changes in traffic behavior (e.g., hostile occurrences) may result in misleading alarms. This work uses fuzzy logic to improve decision-making in anomaly identification.

4.3 Algorithms Based on Genetic Information

In search heuristics, genetic algorithms use crossover, inheritance, mutation, and selection to replicate biological processes. As a result, evolutionary algorithms can classify data and select the best detection settings [30].

Ming. Y [31] suggested utilizing a genetic algorithm and KNN for feature selection and weighting. Weighted training phase features were investigated. DoS assaults were employed to put systems through their paces.

4.4 Neural Networks

Neural networks excel at extrapolating from erroneous data. This ability to generalize enables the recognition of previously unknown patterns that do not fit the previously specified patterns of input. Because the ideal intrusion detector should be able to detect both known and unknown threats, neural networks are a viable anomaly detection technology [32].

Chandrika Palagiri [33] demonstrated an NIDS based on an artificial neural network. Both employed MLP and SOFM. All abnormalities were discovered, albeit a substantial false positive rate was concerning. False positives accounted for 76% of all abnormalities. According to the study, the significant false positive rate is due to a lack of training data for each assault type. The program was fast enough to allow for real-time categorisation.

4.5 Bayesian Networks

Bayesian networks represent the probabilistic connections between model variables. Because it can encode relationships between variables, predict outcomes, and factor in existing knowledge and data, intrusion detection frequently employs this methodology with statistical methods [34].

Aldes et al. [35] created naïve Bayesian networks2 based on traffic spikes for anomaly and intrusion detection. Their technology, which has been implemented into EMERALD [36], has the ability to detect spread attacks where individual attack sessions are not suspicious enough to raise an alarm. However, there are some downsides to this strategy.

4.6 Statistical-Based Anomaly Detection Techniques

Statistical anomaly detection systems analyze people's behavior and generate profiles. Profile components such as activity level measures, audit record distribution, categorical metrics, and ordinal measurements (such as CPU utilization) are common. Every person has both a current and a history profile. Using an abnormality function of all measurements within the profile, the intrusion detection system compares the current profile to the stored profile. As system or network events, such as audit log entries, arriving packets, and so on, are analyzed, it changes the current profile, resulting in an anomaly score (showing the degree of irregularity for the given event). If the anomaly score reaches a certain threshold, the intrusion detection system will raise an alarm [37].

There are advantages to detecting statistical anomalies. Such anomaly detection systems do not require previous knowledge of security weaknesses or assaults. As a

result, such systems may detect zero-day attacks. Long-term detrimental activities that forecast DoS attacks may also be detected using statistical approaches. Portscans are performed often. Portscan distributions differ significantly from traffic distributions, particularly when a bundle contains unusual attributes (for example, a designed packet). As a result, because long-term port scans are atypical, they will be reported [38].

Statistical anomaly detection, on the other hand, has limitations. Skilled attackers may be able to cause statistical anomaly detection systems to ignore unexpected activity. It is also challenging to limit false positives and negatives. Statistical approaches require reliable distributions and cannot capture all behaviors. Most statistical anomaly detection methods assume a semi-stationary process, which differs from most of the data that anomaly detection systems manage [39].

Haystack [40] was the first statistical anomaly-based intrusion detection system. The characteristics of the system were seen as uncorrelated Gaussian random variables, and anomaly detection was employed both individually and collectively. Haystack's parameter settings were standard. If a feature was out of range, the subject's score rose. Under the assumption of independent attributes, score probability distributions were created. A warning was issued because of the high scores. User groups and profiles were retained by Haystack. Users who have not yet been acknowledged were given a new profile with limited access based on their group membership. Malicious usage, data leakage, denial-of-service assaults, and unauthorized user access were all discovered. Haystack was only available offline. Despite the necessity for high-performance systems, statistical studies for real-time intrusion detection failed. Second, because profiles were so critical, system administrators need assistance in spotting intrusive activity indications. The majority of DDoS attacks and network traffic anomalies are identified using statistical techniques such as entropy, correlation, and covariance. Network distinguishing features can be retrieved from traffic recordings or individual packets using statistical techniques.

Statistical or correlation approaches are used in network traffic analysis. They detect network flow attacks. These methods are known as teacher-assisted machine learning.

5 Performance Evaluation

Classification score, inference time, model size, and power consumption were used to evaluate ML model candidates. Ten times, performance metrics were measured. This procedure ensured that extraneous factors had no effect on the consistency of the results.

F-score metrics from categorization reports were used to evaluate performance. The F-score, also known as the F-measure, is the harmonic mean of recall scores and accuracy from the expected label to the actual label. Precision is calculated by dividing true positives (tp) by positive classifications (including false positives) (f p). True positives divided by all positives, including false negatives (f n), equals recall. In binary classification, recall equals sensitivity. F_β is calculated as per Eq. (1):

$$F_\beta - (1+\beta)^2 * \left(\frac{Precision * Recall}{(\beta^2 * Precision) + Recall}\right) \tag{1}$$

False negatives are important for future onboard implementation, and the F1-score ($F_\beta = 1$) has the fewest. F-score measurements were used to assess the dataset's binary

classification and imbalance class problem. The inference time was directly tracked via the Jupyter Notebook package. It quantifies the amount of time it takes the CPU to process instructions. Pickle calculated model size from a Python object translated to a byte stream. The Raspberry Pi experiment looked into power consumption during preprocessing and model inference.

5.1 Rule-Based Anomaly Detection Techniques

Anomaly detection rule-based solutions employ a preset set of rules to identify "typical" human, network, or computer system behavior. These rules look for any irregularities that might suggest a breach by comparing the audit data to a set of specified state change scenarios for penetration.

5.1.1 Expert System

Experts emerge from rule-based systems. The current state of the system is represented by a knowledge base composed of information and rule bases. Audit logs and system activity data are part of a fact base. The rule base contains rules that describe common incursion strategies and other common scenarios. A binding between rules and observable data is created when the pattern of a rule's antecedent conforms to the claim. After binding, all of the pattern occurrences in the rule set must bind in order for the analysis to proceed and related variables to match.

The most well-known rule-based anomaly detection systems are expert systems, which trigger when unusual behaviour is detected in parsed audit data. If observed behavior deviates considerably from predicted behavior after identifying anticipated behaviors by individuals, teams, remote hosts, and intended systems, expert system rules are activated. [41].

The "Next-generation Intrusion Detection Expert System (NIDES)" from SRI International combines a new statistical anomaly detection technology with an expert knowledge system that records typical kinds of intrusion. NIDES considers user-related CPU and file use. A chi-squared statistic is used to find discrepancies between the system's long-term profile and its usage rate or intensity characteristics. Instead of tables, the system was trained to use empirical distributions. Outliers can also be identified by an expert system.

S. Owens et al. proposed a fuzzy set-based Expert system capable of detecting and adapting to intrusions. The membership of fuzzy sets varies. This theory develops a class function for assessing incremental set membership in the real unit space 0–1. Because of its significant uncertainty and ambiguity, this approach may be used in an anomaly detection system and adapt to diverse threats [42].

These rule-based systems place too much confidence in network management and are unable to adapt to changing network circumstances. To solve this issue, T.D. Ndousse and T. Okuda develop a fuzzy cognitive mapping expert system model (FCM). Arcs represent the fault propagation paradigm, whereas nodes represent controlled items like network nodes. FCM simulates the spread and interplay of network faults.

5.1.2 Model-Based

Model-based approaches go beyond rule-based systems in which audit data is applied to expert rules. Anomaly detectors detect unexpected activity by using a precomputed model of typical operation. Flexibility enables more data processing. When the system predicts the intruder's next moves, infiltration efforts may be better explained. It takes a lot of effort to develop a model that balances efficacy and efficiency.

Researchers employed numerous models to define the baseline behavior of the observed system. Model-based techniques detect abnormalities by comparing observed data to the baseline behavior of the system. Researchers employ data mining, neural networks, pattern matching, and prediction models. [43].

5.2 Comparative Analysis

Over the previous two years, anomaly detection approaches for Fog Computing Devices have dramatically improved in accuracy. [51] In 2020, the average accuracy of the methods examined was approximately 70%, but by 2022, it had climbed to 87%. This large increase in accuracy implies that research efforts have resulted in the development of more accurate ways for identifying abnormalities in fog computing devices. In terms of scalability, the outcomes of Anomaly Detection Techniques for Fog Computing Devices have been mixed [52]. In 2020, most strategies were meant to function on small data sets, but by 2022, the emphasis had turned to larger data sets, resulting in a decrease in scalability. However, the general scalability of the methodologies has substantially increased, with the average accuracy increasing from around 50% in 2020 to around 70% in 2022 [53]. This indicates that researchers have significantly developed more scalable and efficient anomaly detection techniques for fog computing devices.

6 Challenges and Future Directions

Before fog computing can be deployed, several challenges must be overcomed. Fog computing presents the following challenges:

A. Fog System SLA: For the time being, fog systems require service level agreements. All current service level agreements (SLAs) for cloud systems apply to fog systems as well. Because a fog system covers several domains, a unique and possibly valuable SLA for fog computing is required.
B. Design of a Bandwidth-Aware Fog System: A significant feature of fog systems is the reduction of primary network bandwidth. It is necessary to conduct research on the best way to utilise bandwidth in a fog computing environment. When more devices are introduced to a network, bandwidth limitations must be kept to a minimum.
C. Several traits Current fog computing systems only account for a narrow subset of qualities, necessitating the development of a new fog design system (QoS, Cost, etc.) and assuming that none of these extra elements would have any influence on the system's functioning. Future planning should include a new scheme design that considers issues such as bandwidth, energy consumption, and cost [49].

D. Scalability of Fog Schemes: The present fog computing methodologies and algorithms are incapable of dealing with the huge scale of IoT networks. For fog systems to be deployed alongside IoT networks, designers must prioritize scalability in their algorithms. For scalable solution, online task offloading is required.
E. Fog Resource Monitoring: When a fog node has several users, it's critical to monitor how it's being used. Future work will need to concentrate on creating a fog-resource monitoring method that allows several operators to use it at the same time.
F. Green Fog Computing: To improve the energy consumption of fog systems, researchers studied computer energy offloading and energy efficiency management of mobility [50].
G. Federated Reinforcement Learning (FRL) can be a promising approach for anomaly detection in fog computing devices, especially when privacy and data decentralization are significant concerns. Anomaly detection is crucial in fog computing to ensure the reliability, security, and efficient operation of the distributed fog environment.

Future research should focus on scalable solutions, resource monitoring, and Federated Reinforcement Learning for anomaly detection in fog systems.

7 Conclusion

According to the findings of this systematic research, anomaly detection algorithms for fog computing devices have the ability to create a safe and robust network for IoT applications. This research identified a variety of anomaly detection strategies, including as statistical-based methods, machine learning, deep learning, and hybrid approaches, that may be adapted and applied to the specific needs of fog computing devices. In addition, the research highlighted some of the important problems and unresolved concerns in the realm of anomaly detection for fog computing devices. This paper gives a thorough overview of the present state of the art in anomaly detection and recommends future research options. In summary, anomaly detection techniques for fog computing devices are essential for maintaining the integrity, security, and performance of fog environments.

These techniques leverage machine learning, real-time processing, and distributed capabilities to identify anomalies and enable prompt responses to emerging threats or operational issues. As fog computing continues to evolve, so will the sophistication of anomaly detection mechanisms to address new challenges in this dynamic computing paradigm. Although further study is needed, present findings demonstrate that anomaly detection techniques may be used to detect and mitigate possible risks to fog computing equipment.

References

1. Abdulkareem, K.H., et al.: A review of fog computing and machine learning: concepts, applications, challenges, and open issues. IEEE Access **7**, 153123–153140 (2019). https://doi.org/10.1109/ACCESS.2019.2947542
2. Askar, S., Jameel Hamad, Z., Wahhab Kareem, S.: Deep learning and fog computing: a review, pp. 197– 208 (2021). https://doi.org/10.5281/zenodo.5222647

3. Rezapour, R., Asghari, P., Javadi, H.H.S., Ghanbari, S.: Security in fog computing: a systematic review on issues, challenges and solutions. Comput. Sci. Rev. **41**, 100421 (2021). https://doi.org/10.1016/j.cosrev.2021.100421
4. Atlam, H.F., Walters, R.J., Wills, G.B.: Fog computing and the internet of things: a review. Big Data Cogn. Comput. **2**, 10 (2018). https://doi.org/10.3390/bdcc2020010
5. Moustafa, N., Hu, J., Slay, J.: A holistic review of network anomaly detection systems: a comprehensive survey. J. Netw. Comput. Appl. **128**(2021), 33–55 (2019). https://doi.org/10.1016/j.jnca.2018.12.006
6. Saranya, T., Sridevi, S., Deisy, C., Chung, T.D., Khan, M.K.A.A.: Performance analysis of machine learning algorithms in intrusion detection system: a review. Procedia Comput. Sci. **171**(2019), 1251–1260 (2020). https://doi.org/10.1016/j.procs.2020.04.133
7. Al Samara, M., Bennis, I., Abouaissa, A., Lorenz, P.: A survey of outlier detection techniques in IoT: review and classification. J. Sens. Actuator Networks **11**(1) (2022). https://doi.org/10.3390/jsan11010004
8. Onah, J.O., Abdulhamid, S.M., Misra, S., Sharma, M.M., Rana, N., Oluranti, J.: Genetic Search Wrapper-Based Naïve Bayes Anomaly Detection Model for Fog Computing Environment, vol. 2. Springer International Publishing (2021). https://doi.org/10.1007/978-3-030-71187-0_127
9. Jaiswal, R., Chakravorty, A., Rong, C.: Distributed fog computing architecture for real-time anomaly detection in smart meter data. In: Proceedings - 2020 IEEE 6th International Conference Big Data Computer Service Application BigDataService 2020, pp. 1–8 (2020). https://doi.org/10.1109/BigDataService49289.2020.00009
10. Onah, J.O., Abdulhamid, S.M., Abdullahi, M., Hassan, I.H., Al-Ghusham, A.: Genetic Algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment. Mach. Learn. Appl. **6**, 100156 (2021). https://doi.org/10.1016/j.mlwa.2021.100156
11. Sharma, D.K., et al.: Anomaly detection framework to prevent DDoS attack in fog empowered IoT networks. Ad Hoc Networks **121**, 102603 (2021). https://doi.org/10.1016/j.adhoc.2021.102603
12. Lawal, M.A., Shaikh, R.A., Hassan, S.R.: An anomaly mitigation framework for iot using fog computing. Electron. **9**(10), 1–24 (2020). https://doi.org/10.3390/electronics9101565
13. Santos, J., Wauters, T., Volckaert, B., de Turck, F.: Fog computing: enabling the management and orchestration of smart city applications in 5G networks. Entropy **20**(1) (2018). https://doi.org/10.3390/e20010004
14. Huč, A., Šalej, J., Trebar, M.: Analysis of machine learning algorithms for anomaly detection on edge devices. Sensors **21**(14) (20210. https://doi.org/10.3390/s21144946
15. Yi, S., Hao, Z., Qin, Z., Li, Q.: Fog computing: platform and applications. In: Proceeding - 3rd Workshop on Hot Topics in Web Systems and Techologies HotWeb 2015, pp. 73–78 (2016). https://doi.org/10.1109/HotWeb.2015.22
16. Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., Amira, A.: Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. Appl. Energy **287**, 116601 (2021). https://doi.org/10.1016/j.apenergy.2021.116601
17. Tang, Z., Chen, Z., Bao, Y., Li, H.: Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. Struct. Control Heal. Monit. **26**(1) (2019). https://doi.org/10.1002/stc.2296
18. Aazam, M., Huh, E.N.: Fog computing and smart gateway based communication for cloud of things. In: Proceedings - 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014, no. August, pp. 464–470 (2014). https://doi.org/10.1109/FiCloud.2014.83

19. Rahul, M., Alhumiany, H., Muntjir, M., Alhumyani, H.A.: An Analysis of Internet of Things(IoT): Novel Architectures, Modern Applications, Security Aspects and Future Scope with Latest Case Studies Task Scheduling in Cloud Computing View project Internet of Things View project An Analysis of Internet of Things, vol. 6, no. June, pp. 422–448 (2017)
20. Basora, L., Olive, X., Dubot, T.: Recent advances in anomaly detection methods applied to aviation. Aerospace **6**(11) (2019). https://doi.org/10.3390/aerospace6110117
21. Bulusu, S., Kailkhura, B., Li, B., Varshney, P.K., Song, D.: Anomalous example detection in deep learning: a survey. IEEE Access **8**(Ml), 132330–132347 (2020). https://doi.org/10.1109/ACCESS.2020.3010274
22. Ma, X., et al.: A comprehensive survey on graph anomaly detection with deep learning. IEEE Trans. Knowl. Data Eng., no. August (2021). https://doi.org/10.1109/TKDE.2021.3118815
23. Pacheco, J., Hariri, S.: Anomaly behavior analysis for IoT sensors. Trans. Emerg. Telecommun. Technol. **29**(4), 1–15 (2018). https://doi.org/10.1002/ett.3188
24. Yaseen, Q., Albalas, F., Jararwah, Y., Al-Ayyoub, M.: Leveraging fog computing and software defined systems for selective forwarding attacks detection in mobile wireless sensor networks. Trans. Emerg. Telecommun. Technol. **29**(4), 1–13 (2018). https://doi.org/10.1002/ett.3183
25. Xiaonan, S., Wolfgang, W.: Computer Science (2008)
26. Elrawy, M.F., Awad, A.I., Hamed, H.F.A.: Intrusion detection systems for IoT-based smart environments: a survey. J. Cloud Comput. **7**(1), 1–20 (2018). https://doi.org/10.1186/s13677-018-0123-6
27. Su, M.: Expert systems with applications real-time anomaly detection systems for denial-of-service attacks by weighted k-nearest-neighbor classifiers. Expert Syst. Appl. **38**(4), 3492–3498 (2011). https://doi.org/10.1016/j.eswa.2010.08.137
28. Badidi, E., Mahrez, Z., Sabir, E.: Fog computing for smart cities' big data management and analytics: a review. Futur. Internet **12**(11), 1–29 (2020). https://doi.org/10.3390/fi12110190
29. Rekha, G., Malik, S., Tyagi, A.K., Nair, M.M.: Intrusion detection in cyber security: role of machine learning and data mining in cyber security, vol. 5, no. 3, pp. 72–81 (2020)
30. Pareek, K., Tiwari, P.K., Bhatnagar, V.: Fog Computing in Healthcare: a Review. IOP Conf. Ser. Mater. Sci. Eng. **1099**(1), 012025 (2021). https://doi.org/10.1088/1757-899x/1099/1/012025
31. Valdes, A., Skinner, K.: Adaptive, model-based monitoring for cyber attack detection. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 1907, pp. 80–93 (2000) 10.1007/3- 540-39945-3_6
32. Porras, P.A., Neuman, P.G.: EMERALD: event monitoring enabling responses to anomalous live disturbances. In: 9th ACM Conference on Computer and Communications Security, pp. 245–254 (2001)
33. Samann, F.E.F., Abdulazeez, A.M., Askar, S.: Fog computing based on machine learning: a review. Int. J. Interact. Mob. Technol. **15**(12), 21–46 (2021). https://doi.org/10.3991/ijim.v15i12.21313
34. Haji, S.H., Ameen, S.Y.: Attack and anomaly detection in IoT networks using machine learning techniques: a review. Asian J. Res. Comput. Sci. 30–46 (2021). https://doi.org/10.9734/ajrcos/2021/v9i230218
35. Raut, R., Sawant, R., Madbushi, S.: A review of security issues in cognitive radio. Cogn. Radio. 137–151 (2020). https://doi.org/10.1201/9780429353109-8
36. Smasha, S.E.: An_intrusion_detection_system.pdf. (1988)
37. Montoya-Munoz, A.I., Rendon, O.M.C.: An approach based on fog computing for providing reliability in iot data collection: a case study in a colombian coffee smart farm. Appl. Sci. **10**(24), 1–16 (2020). https://doi.org/10.3390/app10248904
38. Mahmud, R., Ramamohanarao, K., Buyya, R.: Application management in fog computing environments: a taxonomy, review and future directions. ACM Comput. Surv. **53**(4) (2020). https://doi.org/10.1145/3403955

39. Saurabh, Dhanaraj, R.K.: A review paper on fog computing paradigm to solve problems and challenges during integration of cloud with IoT. J. Phys. Conf. Ser. **2007**(1) (2021). https://doi.org/10.1088/1742-6596/2007/1/012017
40. Qin, B., Tang, H., Chen, H., Cui, L., Liu, J., Yu, X.: Review on big data application of medical system based on fog computing and IoT technology. J. Phys. Conf. Ser. **1423**(1), 1–9 (2019). https://doi.org/10.1088/1742-6596/1423/1/012030
41. Abdullahi, M., et al.: Detecting cybersecurity attacks in internet of things using artificial intelligence methods: a systematic literature review. Electron **11**(2), 1–27 (2022). https://doi.org/10.3390/electronics11020198
42. Aazam, M., Zeadally, S., Harras, K.A.: Offloading in fog computing for IoT: review, enabling technologies, and research opportunities. Futur. Gener. Comput. Syst. **87**, 278–289 (2018)
43. Labiod, Y., Amara Korba, A., Ghoualmi, N.: Fog computing-based intrusion detection architecture to protect iot networks. Wirel. Pers. Commun. **125**(1), 231–259 (2022)
44. Mahmud, R., Pallewatta, S., Goudarzi, M., Buyya, R.: IFogSim2: an extended iFogSim simulator for mobility, clustering, and microservice management in edge and fog computing environments. J. Syst. Softw. **190**, 111351 (2022)
45. Saurabh, D.R.K.: Enhance QoS with fog computing based on sigmoid NN clustering and entropy-based scheduling. Multimed. Tools Appl. (2023). https://doi.org/10.1007/s11042-023-15685-3

Supermarket Sales Prediction Using Linear Regression: A Case Study Approach

Anuja Bokhare[1](✉), Ojas Pawaskar[2], Kriti Bhatia[3], and Tirupathi Mandala Reddy[4]

[1] School of Computer Science and Engineering, Dr. Vishwanath Karad MIT World Peace University, Kothrud, Pune 411038, MH, India
anuja.bokhare@gmail.com
[2] Intellimation.ai, Pune, Maharashtra, India
[3] Infocusp Innovations, Pune, Maharashtra, India
[4] The Intelli, LLC, Edison, NJ, USA

Abstract. Machine learning has been a subject going through exceptional examination across various enterprises and luckily, organizations are getting step by step more mindful of the different ways to deal with to take care of their issues. The customary methodology of deals and showcasing objectives at this point doesn't help the organizations, to adapt up to the speed of the serious market, as they are done with no experience with clients' buying designs. Major changes can be found in the area of deals and promoting because of Machine Learning headways. In this paper, the business dataset of supermarkets and attempt to examine how it can create later on and furthermore what are the fundamental advances that should have been taken for its improvement and consumer loyalty. This will be continued in a step-by-step investigation report from the recorded information and its implementation in the future. The study includes an extensive examination of sales prediction, employing Machine Learning techniques such as Linear Regression, as detailed in the research.

Keywords: Machine Learning · Supermarket · Model · Linear Regression · Sales

1 Introduction

In the present current world, colossal retail plazas like large shopping centers and stores are recording information identified with deals of things or items with their different needy or free factors as a significant advance to be useful in the expectation of future requests and stock or management. The dataset worked with different reliant and autonomous factors is a composite type of item attributes, information accumulated by methods for a client, and furthermore information identified with inventory management. The data is from that point refined to get precise expectations and assemble new just as intriguing results that shed new light on our insight concerning the errand's information. This would then be able to additionally be utilized for anticipating future sales by methods for utilizing machine learning algorithms like the linear regression model.

As on date, progressively the data that's being generated is increasing and there is a need for this large volume of unprocessed data to be analyzed accurately to provide

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 197–208, 2026.
https://doi.org/10.1007/978-3-031-98138-8_16

useful information and insights based on current standard requirements. With the last two decades contributing immensely towards the growth and evolution of Artificial Intelligence (AI), Machine learning (ML) is propelling forward at a rapid pace in its evolutionary cycle. ML is an actively growing and predominant field in the IT sector, although hidden, it's quite wide spread. With the given potential technological growth spurt the understanding and utilization of data to provide insights based on the current requirements is plentiful. ML discusses both supervised and unsupervised tasks and in case of Knowledge discovery, Classification problems contribute and account for a lot. Various resources can be generated and regression can be implemented to make accurate predictions, and focus on converting the system to be more self-sufficient. Data can give a lot of knowledge and information by using the right statistical and probabilistic tools. Sampling distributions are used as conceptual keys in statistic inferencing.

ML can take many forms. In this paper various implementations of ML and the kind of data they affect are scrutinized. The statement that has been discussed through this paper has been presented in definitive manner. The methodology is well explained and the prediction results have been scrutinized after implantations and observation.

Based on a given set of independent variables are used to predict a continuous or dependent variable, this technique is known as linear regression. This is said to be parametric as different assumptions are made on basis of the data set.

2 Previous Study

In the study conducted by Rising Odegua [1] investigates forecasting sales for a supermarket chainstore called "Chukwudi Supermarkets". Author applies machine learning algorithm for testing namely Random forest, Gradient Boosting and K-Nearest Neighbor. Random forest gives good prediction among all. In the study conducted by Oliver Vornberger [2], have developed a model, a multilayer perceptron which is feedforward to accept the advertising campaigns, prices and times series of sales. This model was trained using the back propagation model.this was compared with various topologies and several training parameters, in order to enhance the same the author used additional price with information on advertisements. Authors Mansi Panjwani et al. [3] designed a best-suited predictive model that would be suggested for a company to forecast sales trends. The study gives us a summary of the various ML techniques on the basis of how accurate the predictions are. The authors have mainly used three prominent models and used past and present sales data to train and test the classification model. By comparing the models, more accurate forecasting could be presented. When the results of the single classifiers with the results from general model with 83.33%.

Aurthors, Robert Siwerz and Christopher Dahlen [4] conducted the study to match three machine learning models for predicting the sales, specifically for the food industry. The study scrutinizes the methods Radial Basis Function Network (RBFN), Support Vector Machine (SVM) and Multilayer Perceptron (MLP) based on their performance measures and how accurate their results were. Mean Average Percentage Error (MAPE) and Root Mean Squared Error (RMSE) were used as performance measure for the study. In order to figure out if there was any differences between the models they utilized repeated measure analysis of variance (ANOVA). The study concluded that there was a statistically observable difference between the methods mentioned.

In this research author, Judi Sekban [5] has used four different algorithms, a stacking ensemble technique, and a specific approach to feature selection to develop models. The results of feature selection achieved an improvement of 3(%) in the R2 value compared to the first configuration of the models, and by 2(%) compared to the tuned models. In other words, author could predict sales with a better level of accuracy using only five features. Author, Kyawt Kyawt San, [6] has applied regression analysis to estimate future sales values or values of a variable using information of other features. In the study, San used linear regression, random forest regression and K-Nearest Neighbors (KNN) regression to experiment with using Myanmar supermarket sales dataset. The main objective of the research is to relate performance of regression analysis among these regressors. According to the research, the linear regression model achieves the best among these regression models. The paper also intends to research all three regressors and analyze the best analyzer for supermarket sales data analysis. Authors, Ana Lucia Silva and Margarida G. M. S. Cardoso [7] have studied how data reduction in the case of store attributes like visibility and accessibility can be done by using Principal components analysis. For this the authors have taken into consideration factors that affects the sale like competitors or availability of transportation and have used regression tree model to overcome the same. Authors, Kavya Ramesh et al., [8] used analytical techniques like Generalized Linear Model, Linear Regression, and Artificial Neural Networks were verified in order to come up with correct result and to identify the most suitable technique with best predicting results. Through real time training and testing of data, a clear pattern could be obtained which helps in the analyzing the performance and predicting the sustainability of the business. Understanding the sales pattern will also help predict the product movement which in turn can be utilized to bring in the product with accurate quantity thereby minimizing the wastage.

Authors, Heramb Kadam et al., used [9] multiple learning algorithms such as multiple linear regression and random forests. Random forests provide a solution for the decision trees' habit of overfitting. Hence, authors have proposed a software tool for forecasting future sales volume based on the historical sales data collected. Using this tool, the accuracy of prediction for multiple linear regressions and random forests could be determined. Authors, Daljeet Kaur and Jagroop Kaur, [10] in their paper have discussed how the economy is propelling fast and how this development has raised a need for utilization of data available for gathering insights for retail purposes using knowledge mining especially in case of large-scale super markets. For the same reason, it has become necessary for to develop and institute various data mining techniques and models to properly use data and aid in the process of decision making.

Authors, Vitaliy Buyar and Amal Abdel-Raouf, [11] presented, a convolutional neural networks-based model. The model was used to predict future sales for a pharmaceutical company using its real large-scale sales data. The prediction results were evaluated based on the mean absolute error and mean absolute percent error metrics, which were used to determine the accuracy and show the effectiveness of our model. Authors, Shenjia Ji et al., have [12] proposed a prediction model in their study which combines ARIMA model with BP neural network, in order to overcome the limitations that are faced when used ARIMA model. Simply using the single model would lead to low efficiency when predicting long term economic activity especially taking the future forecast into account.

This proposed model has shown higher accuracy in comparison to the single model. This increased efficiency is due to combining the describing ability of stationary time series developed from the statistic model and the neural network model's strong fitting ability of non-linear variable.In the research paper given by Michael Giering, [13] the focus is on enhancing the sale in case of a major retailer and to enhance the forecasting of sale by utilizing four distinct types of data and taking their distinctive advantage. The authors used retailer defined types of customer and data at the level of the store which was categorized by items. The model that was built was utilized as an analysis tool based on outliers as well as the final product. In case of items of interest there was 1.5–5 times greater accuracy for prediction, based on r-squared error statistics. This can be utilized in real life scenarios such as new store locations stocking.

Authors, You Lingxian et al., [14] in their paper ways to utilize various historical information to aid in the prediction of retail dynamics online. To achieve this they have developed an architecture which integrates unsupervised learning of data and K-mean clustering while applying long short term memory of an advanced artificial intelligence model. It was found out that in order to efficiently allocate resources while also reducing the cost, when it came to working under a 20-h time window it is better to use retail activity which occurred during a 10-h time slot for better understanding of online dynamics.

Authors, Hai Rong et al., [15] have presented how micro competitions can be analyzed better and retail store sales predictions could be optimized by using simulation techniques by developing an analysis framework based on the gravity model and GIS platform but is much more elaborate in evaluation of the impact of various factors which may affect the sale. This model also takes into account the effects of new stores parallel to the existing stores that are competing and provides an analysis of the same, so that new sites for the purpose of business development could be identified. Sales of selected site of new stores is also estimated by the same.

In the paper proposed by Patrick Meulstee and Mykola Pechenizkiy, [16] mainly focuses on the operations of business and how the sales for various products can be enhanced by developing new feature groups. In the study observed that when handling the seasonal changes for several products it is better to group them with a feature set which encompasses similar sales pattern which objects together. The feature takes into consideration the extreme seasonal changes (like taking holidays and weather) and have a lot of space for improvement must also be taken into account. Linear regression was used in this study for supermarket sales prediction. Study explains the establishment of a neural network prediction algorithm for a large German supermarket chain. [17] the method uses BPP neural networks to estimate weekly consumer interest in particular items, considering elements such as cost, marketing initiatives, and vacation. The authors analyze the artificial neural network method's execution to two standard algorithms for forecasting and finding that it surpasses both in terms of forecasting efficiency. The technique was developed to analyze data effectively throughout the series and to utilize the same network framework over all 20 profit time series. The analysis finalizes with a discussion about forthcoming work, which will centre on implementing the forecasting tool into the more general flow of data process for the supermarket and building further versatile transformation methods into consideration. The study proposes [18] a method

for foreseeing supermarket sales that uses periodic forecasting to figure out reserve levels. In order to produce accurate range forecasts and normalize the distribution function of cumulative values, this approach leverages continuously categorized quantile regression analysis. The findings from experiments show that experimental strategies surpass outdated techniques, particularly regarding robust point forecasting. The machine learning techniques can improve sales projections for certain retail products, benefiting the economy and environment. Using two years of sales, calendar events, and meteorological data, the thesis compared Facebook Prophet, LSTM, XGBoost, and ARI-MAX. XGBoost and LSTM models performed best overall, while Facebook Prophet performed best for holidays [19]. The weather did not improve models, and results show the optimum model depends on the forecast time frame and objective. Big Marts, run-of-the-mill supermarkets, monitor each item's sales data with the objective of forecasting potential consumer demand and revise inventory control. By mining the data from the data warehouse, anomalies and broad trends are discovered. The predictions can be utilized by retailers like Big Mart to predict future sales volume using a variety of machine learning approaches. For projecting the sales of a company like Big-Mart, a predictive model was created utilizing XGboost and various other regression techniques such as Linear, Polynomial and Ridge regression. The model was discovered to have a better performance as compared to the others [20].

To ensure sustainable competition, businesses must take proactive measures for the upcoming period, and accurate sales quantity and sales income forecasting play a crucial role in this regard. For expanding industries like the grocery retailing sector, it is particularly crucial. The retailing of groceries in Turkey is developing quickly. The purpose of this study [21] is to predict sales revenue for the Turkish grocery retailing business using artificial neural networks to take into account marketing expenses, gross profit, and competitors' gross profit. Because of their ability to recognize patterns efficiently and machine learning, artificial neural networks are selected. The ANN approach is used to predict future sales income. The results show that the actual data and the anticipated data have a high degree of similarity. This paper discusses the accessibility of artificial neural networks for forecasting a supermarket's energy utilization. The analysis compares the statistical outcomes of ANNs with the standard multilevel regression approaches and delivers first indications of moderately accurate 1/2-hourly figures of the electrical energy consumption in supermarkets [22]. The study's authors establish the autonomous key inputs or variables enabled through internet forecasting in food distributors' shops for cooling and heating, and air conditioning diagnostics, process control, maximizing efficiency, and energy optimization. Based on the study's findings, the ANN method proposes a general tool for visualizing relations within inputs and outputs which needs a lesser amount of expertise and creative thinking than standard forecasting strategies.

Current study focuses on applying machine learning algorithms for performing supermarket prediction. Section 3 explains the proposed methodology of the same. Section 4 discuss about the dataset and experiment discussion. Section 5 illustrate the result analysis and Sect. 6 and 7 devoted for discussion and conclusion of the research study.

3 Methodology

The current study makes use of supervised learning. A subset of machine learning and artificial intelligence, supervised learning is often referred to as supervised machine learning. Its use of labelled datasets to train algorithms to classify information or accurately predict results and outputs defines it. The model adjusts its loads as the information document is fed into it through a reinforcement learning process, ensuring that the model has been fitted correctly. Identifying spam in a different folder from your inbox, for example, is one of the many certifiable concerns that supervised learning helps **organizations** address at **scale. In** supervised learning, models are shown to provide the desired output using a training set. This training dataset has both valid inputs and outputs, allowing the model to learn over time. Using the loss function, the method calculates its precision and accuracy and iterates until the error has been sufficiently reduced. Furthermore, in terms of algorithms, simple linear regression is implemented during the study. Figure 1 describe the prediction model proposed.

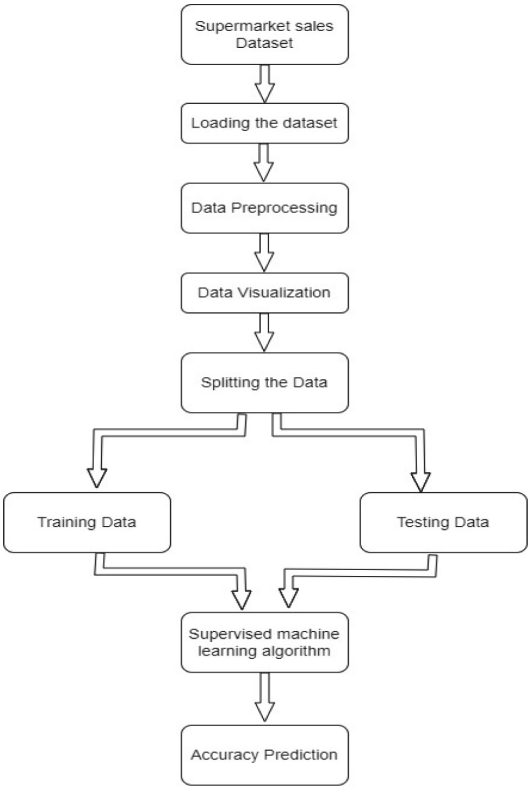


Fig. 1. Flow diagram of the Model

4 Dataset and Experiment Discussion

The dataset for this project has been taken from Kaggle (https://www.kaggle.com/). This dataset consists of data pertaining to supermarket has 17 different attributes (columns) with 1000 records (rows).

The initials step in the process of building the model is to first load in the dataset and then understand it. For the process of model building, there are many ways and to list a few:

- Prediction
- Classification
- Clustering

Then came the splitting of the Training set from the Testing set. Figure 2 and Fig. 3 shows that as the unit price increases, the gross percent spreads on both sides of the line, i.e., the rate of variance of gross percent increases with unit price. So projecting graphs further, it can determine a unit price in accordance with the variance.

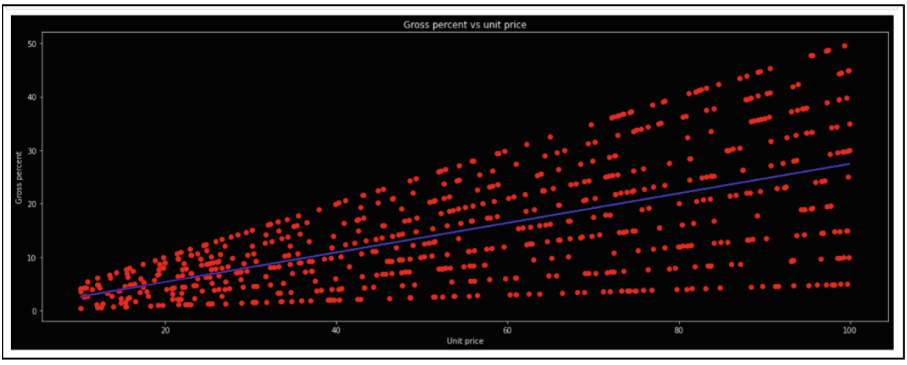


Fig. 2. Gross percent vs Unit Price

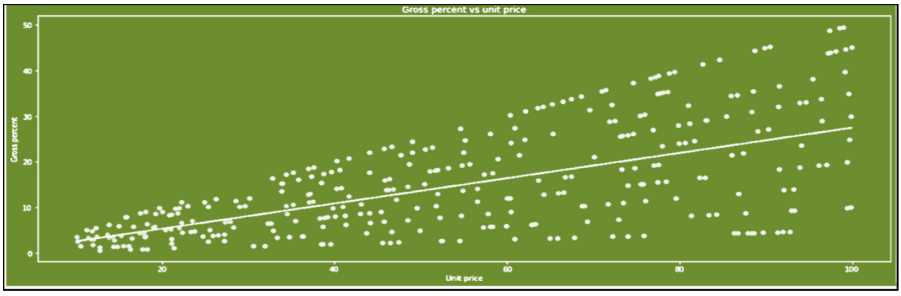


Fig. 3. Gross percent vs Unit Price (*varied*)

5 Result Analysis

To better understand the data and model, various intricate and carefully studied visualizations are developed for better understanding of the data at hand.

The visual representation shown in Fig. 4 is a group of pie charts discussing the different ways each model is partitioned in. These are namely, Supermarket branches, Cities of Shops, Payment modes and Membership/Customer Type. The Supermarket pie chart has three different branches split evenly which are A, B and C. The second one to the right showcases the 3 cities where the shops are located. Following the one in the bottom left corner shows the slices of the types of payment modes which are accepted and used more often for any purchases by customers which are credit cards, E-wallet and finally cash. The fourth pie chart tells us about the two different types of customers that visit the supermarkets. They are split into ones that are members and joined a membership offered by the supermarket and the other which are regular come and go customers otherwise classified as "Normal".

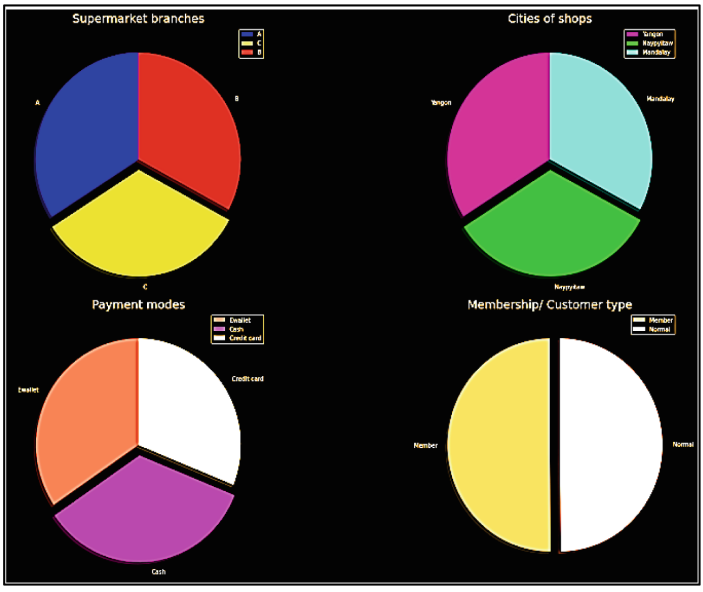


Fig. 4. Categorical Classification

Figure 5 consist of countplots of the three branches the supermarket is divided into. From the visual representation we can identify that branch C flourishes more than its sister branches A and B in terms of profit. With a gross income of over 15% Branch C saw the sales of its products from its supermarket high in demand with more customer revenue that the other two branches. It can clearly realize that branch C has the highest revenue amongst the other three centers.

Figure 6 find out which city would it be most profitable to open a shop. Figure 6 is a visual representation of the cities in the country of Myanmar in which these supermarkets

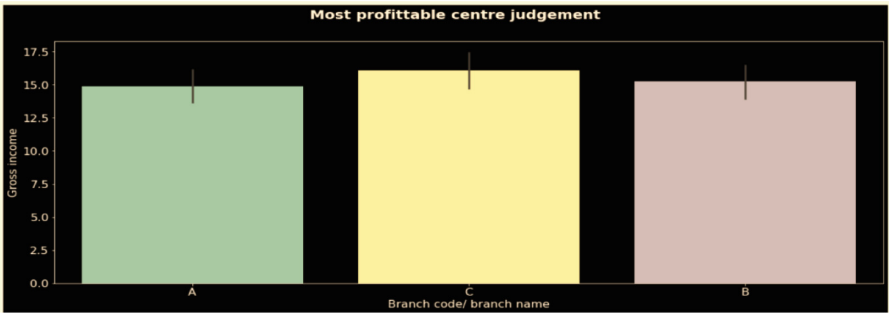


Fig. 5. Most Profitable Centre

run their business. The subplot consists of countplots of the three cities namely, Yangon, Naypitaw and Mandalay. From the three cities we can clearly see that all three cities perform well in terms of gross income but from further analysis the city with the most profit gained is the city Naypitaw with a gross income of over 15%.

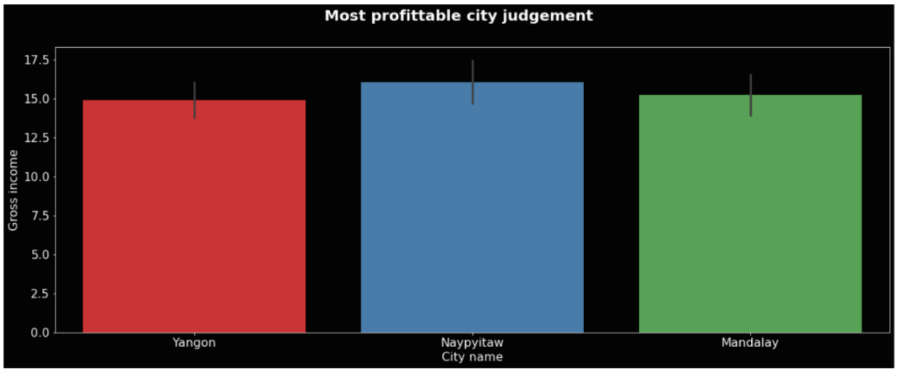


Fig. 6. Most Profitable City

Another visual representation would be of the "Rating" shown in Fig. 7. Figure 7 shows us the rating scale of customers with their respective counts on the products purchased from their various supermarkets. The maximum ratings are in 6.0 which denotes that satisfaction level of the super-market commodities and services are above average and is quite good (considering some spam reviews as well). There are many people who have given a rating of 9.7 which bodes well with the overall performance of the supermarket stores along with the products at hand.

Maximum ratings are in 6.0 which denotes that satisfaction level of the super-market commodities and services are above average and is quite good (considering some spam reviews as well). Many people have also voted for 9.7 as a rating. Table 1 depicts the various performance metrics for the prediction model.

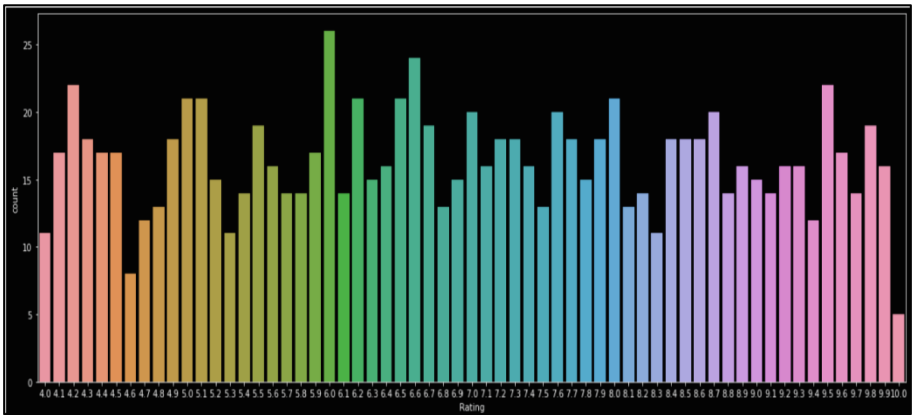


Fig. 7. Customer Satisfaction Scale

Table 1. Performance Metrics of the proposed model

Performance Metric Name	Value
Precision	67%
Recall	80%
F1 – score	73%
Accuracy	71%

6 Discussion

The projected model here, after performing several trainings and testing on it gave a calculated accuracy of 71%. This is likely because of the time constraint put on this study. Here what was primarily down prioritized was finding the optimal parameter settings for the model. One factor that could highlight the limitations of the study is the lack of in-depth knowledge in determining the optimal parameter settings. The study's findings may be affected by this deficiency, potentially impacting the accuracy and robustness of the results. Indeed, possessing more comprehensive and detailed information about the hyperparameters would enhance the validity and reliability of the study's results when comparing them to existing literature. A deeper understanding of the hyperparameter settings would enable researchers to make more accurate comparisons, leading to more meaningful insights and potential advancements in the field. The lack of data or not having sufficient amounts of data does not help with the information while statistically comparing the results between machine learning methods. Therefore, a loss of generality has to be considered. The study's limitation to only one store is primarily attributed to two significant factors: time constraints and limited data availability. Expanding the scope of the study to include multiple stores could have yielded more generalizable results. Sacrificing either of these factors might have allowed for a broader and more representative

analysis, potentially leading to more comprehensive conclusions applicable to a wider range of scenarios or retail settings.

7 Conclusion

The goal of this research is to develop a system for making forecasts for supermarkets. Because of its adaptability to changing environments, this advanced technology will reduce stock-keeping costs. The following findings came from the study's data analysis across numerous variables and processing:

- The gross percent spreads with the unit price, i.e. for a unit price of higher value, there are a wider range of gross percentages of the commodities in sale.
- Branch C of the supermarket makes the greatest profit
- City of Naypyitaw draws the maximum percentage of the customers
- Most people have voted upto a rating of 6.0 which says that the services and facilities are good for all branches
- Mode of payment doesn't matter at all

Sales forecasting is critical for all businesses, especially large ones, and the process is complicated due to the numerous aspects that must be considered. Supermarket chains always aim to predict sales in order to establish attainable targets and effectively achieve them.

8 Future Scope

In terms of future updates to this work and its scope, the modelling of the input vectors should be enhanced in order to decrease and minimize prediction error. Various different information such as the data that can be collected during the seasons and the holidays will provide a wider perspective of things and will also help in modelling the values of changing prices quantitatively in future.

References

1. Odegau, R.: Applied machine learning for supermarket sales prediction (2020)
2. Vornberger, O.: Short term prediction of sales in supermarkets (1995)
3. Panjwani, M., Ramrakhiani, R., Jumnani, H., Zanwar, K., Hande, R.: Sales prediction system using machine learning (2020)
4. Siwerz, R., Dahlen, C.: Predicting sales in a food store department using machine learning (2017)
5. Sekban, J.: Applying machine learning algorithms in sales prediction (2019)
6. San, K.K.: Performance analysis of regression models using Myanmar sales data (2020)
7. Silva, A.L., Cardoso, M.G.M.S.: Predicting supermarket sales: the use of regression trees (n.d.)
8. Ramesh, K., Mathew, J.J., Hemalatha, N.: Machine learning approach for the predictive analysis of Salesin grocery store (2017)
9. Kadam, H., Shevade, R., Ketkar, D., Rajguru, S.: A forecast for big mart sales based on random forests and multiple linear regression (2018)

10. Kaur, D., Kaur, J.: Data mining in supermarket: a survey (n.d.)
11. Buyar, V., Abdel-Raouf, A.: A convolutional neural networks-based model for sales prediction (2019)
12. Ji, S., Yu, H., Guo, Y., Zhang, Z.: Research on sales forecasting based on ARIMA and BP neural network combined model (2016)
13. Giering, M.: Retail sales prediction and item recommendations using customer demographics at store level (n.d.)
14. You, L., Kou, J., Wang, S.: Online retail sales prediction with integrated framework of K-mean and neural network (2019)
15. Lv, H.R., Bai, X.X., Yin, W.J., Dong, J.: Simulation based sales forecasting on retail small stores (2008)
16. Meulstee, P., Pechenizkiy, M.: Food sales prediction: "if only it knew what we know" (2008)
17. Taylor, J.W.: Forecasting daily supermarket sales using exponentially weighted quantile regression. Eur. J. Oper. Res. **178**(1), 154–167 (2007)
18. Fredén, D., Larsson, H.: Forecasting daily supermarkets sales with machine learning (2020)
19. Ranjitha, P., Spandana, M.: Predictive analysis for big mart sales using machine learning algorithms. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1416–1421 (2021)
20. Dey, S., Ghose, D.: Artificial neural network: an answer to right order quantity (2020)
21. Penpece, D., Elma, E.: Predicting sales revenue by using artificial neural network in grocery retailing industry: a case study in Turkey. Int. J. Trade Econ. Financ. (2014)
22. Datta, D., Tassou, S.A., Marriott, D.: Application of neural networks for the prediction of the energy consumption in a supermarket (1997)

Employees Performance Metrics Using Machine Learning: A Systematic Literature Review Using Prisma Model

Abhilasha Dixit[1](✉), Rashmita Singh[1], Nitin Dixit[2], and Shaifali Garg[1]

[1] Amity Business School, Amity University Madhya Pradesh, Gwalior, India
adixit@gwa.amity.edu
[2] ITM University, Gwalior, Madhya Pradesh, India

Abstract. This study investigates the impact of machine learning on employee performance. Machine learning has been adopted as a metric for assessing employee performance. These indicators have a broader application than only assessments and professional progression. Furthermore, they are critical in facilitating corporate expansion and generating organizational profitability. According to research, precise measures have the potential to greatly improve employee well-being, productivity, and retention inside firms. Machine learning promises unrivaled prospects for people and brands all across the world. This study includes a complete systematic literature review (SLR) on the use of machine learning (ML) in measuring employee performance matrix. The SLR is based on Scopus indexed database where 248 recent papers were chosen from 6363 search results using a PRISMA model-based approach. The review also includes research objectives such as investigating the variety of methods used in machine learning, issues, opportunities, and barriers that are deemed significant machine learning roles, and establishing how machine learning impact employee performance metrics measurement functions and the nature of such impact, as well as a pictorial representation via cluster to justify the methodologies used in addressing the issue.

Keywords: Machine Learning · Employee Performance · PRISMA · Visual representation · Systematic Literature Review

1 Introduction

In the current era of business, characterized by intense competition, firms are increasingly recognizing the utmost importance of effectively evaluating and managing employee performance. The translation of strategies into practical terms is a crucial aspect of modern performance assessment systems (Kamble & Gunasekaran, 2019; Hawkins, M., 2022; Lyons, N., 2022). The standard methods utilized to assess employee performance often demonstrate limits in effectively capturing the complex facets of an individual's contributions and potential for future success. Machine learning, which is a subfield of artificial intelligence, plays a significant role in this particular situation (Cravero et al., 2022). Machine learning has the potential to significantly alter the evaluation,

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 209–218, 2026.
https://doi.org/10.1007/978-3-031-98138-8_17

understanding, and improvement of employee performance by leveraging data analytics and advanced algorithms (Sahinbas, K., 2022). The application of machine learning in the evaluation of employee performance measures involves the utilization of computational techniques to analyze a diverse range of data sources (Nithya & Ilango, 2017).

However, it is crucial to acknowledge that the incorporation of employee performance indicators using machine learning is not without its inherent complications (Wang, et al., 2018). The matters pertaining to privacy concerns, data quality, and the possibility for algorithmic bias are of paramount significance and necessitate scrupulous handling. There are several widely used machine learning tools employed for constructing predictive models and detecting trends in employee performance data (Coates et al., 1992). These tools include Tableau, Power BI, Python with Scikit-Learn, RapidMiner, and others (Choudhury et al., 2020; Shilbayeh & Abonamah, 2021).

Moreover, it is imperative to recognize that while data-driven insights possess significant worth, they should be employed in tandem with, rather than as a replacement for, human discernment and a thorough understanding of individual and group dynamics (Nadler & McGuigan, 2017).

In the contemporary and ever-changing business landscape, there is a growing imperative for enterprises to harness the potential of machine learning to optimize the utilization of employee performance measures (Sharma & Sharma, 2017; Winarno & Widyatmojo; 2019). By using these strategies, firms can effectively strengthen their competitive advantage through the development of a corporate culture that prioritizes ongoing improvement, focused skill development, and efficient talent management. This study aims to analyze existing literature pertaining to the use of machine learning in the assessment of employee performance, including its applications, methodology, and relevant factors to be considered. As the investigation advances, a comprehensive examination of prior research will be conducted. The integration in question possesses the capacity to exert a substantial impact on the future of workforce optimization.

2 Systematic Literature Review: Research Structure Followed by PRISMA Model

To investigate the impact of machine learning on the performance evaluation matrix, we conducted a bibliometric study. Bibliometric analysis has the potential to assess science, scientists, and scientific activity in a systematic, open, and reproducible way (Broadus, 1987; Pehar, 2011; Mody et al., 2021).

We used SCOPUS Elsevier's "Social Sciences" and "Computer Science" indexes to compile a list of articles published on the topic of performance matrices and machine learning over the past decade. To further grasp the significance of machine learning in measuring employee performance, we chose these two Scopus subject areas.

To guarantee that only relevant and high-quality studies were included in the review, the section's primary research centers on topics that are meant to be answered by the inclusion and exclusion criteria. From the PRISMA Protocol's Stage1 (research questions) to its Stage 2 (database searching) to its Stage 3 (eligibility criterion definition) to its Stage 4 (quality criterion definition) to its Stage 5 (screening and data extraction criteria to carry out the research defined while result analysis took place), Fig. 1 shows the many steps involved in a review.

3 Implication of PRISMA Model

The paper begins with a model, then moves on to an examination of Scopus and results, a mapping of research trends based on a close reading of the accompanying bar charts, a discussion of the primary issues posed by the machine learning, and finally a look at some potential solutions.

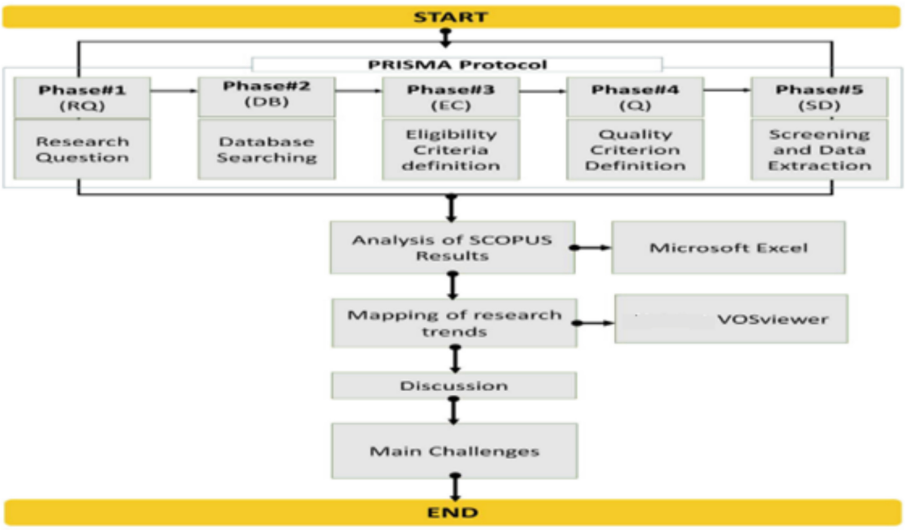


Fig. 1.

With the PRISMA model's guidance, researchers may conduct systematic literature reviews with greater precision, bolstering the trustworthiness of their findings and making them more applicable to future studies and clinical applications.

Stage 1: Formulation of Research Question, Analysis of Data, and Literature Search.

To what extent does machine learning factor into evaluations of staff performance? The answers to these research topics may shed light on the practical use of machine learning methods to the study of performance evaluation.

Stage 2: Identifying Resources via Database Search Scopus is one of the largest academic citation databases in the world, containing articles from thousands of journals,

books, conference proceedings, and scholarly publications across a wide range of fields (Hofmann et al., 2019). It is employing highly targeted search terms, Boolean operators, and granular filters. The research investigation was started with the keywords "Performance Matrix AND Machine Learning *", with Boolean operators "AND", which were processed as: (TITLE-ABS-KEY ("performance matrix" OR "performance measurement" OR "performance appraisal" OR "performance evaluation" AND "machine learning" OR "ML" OR "knowledge engineering") AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SRCTYPE, "j")). Only articles from peer-reviewed national publications were considered for inclusion in this study. The decision was based on the understanding that scholarly articles published in journals are more reliable and rigorous.

Stage 3: Eligibility Criteria Definition Inclusion and exclusion criteria were identified, including documents not related to the machine learning, duplicate documents, or documents written in a language other than English. This was done in accordance with the PRISMA Model, which aims to reduce the potential for bias in systematic reviews.

Stage 4: Documents having a high impact factor, high journal rank, and high cite score that investigate the use of machine learning in performance evaluation constitute the fourth quality criterion.

Stage 5: To guarantee openness and precision in the screening and data extraction at final stage, we followed the PRISMA procedure (preferred reporting items for systematic reviews and meta-analysis). We narrowed the field down to 248 papers from a pool of 6363 by only considering articles written in English that met our criteria for Computer Science, Social Science (depicted in Fig. 2).

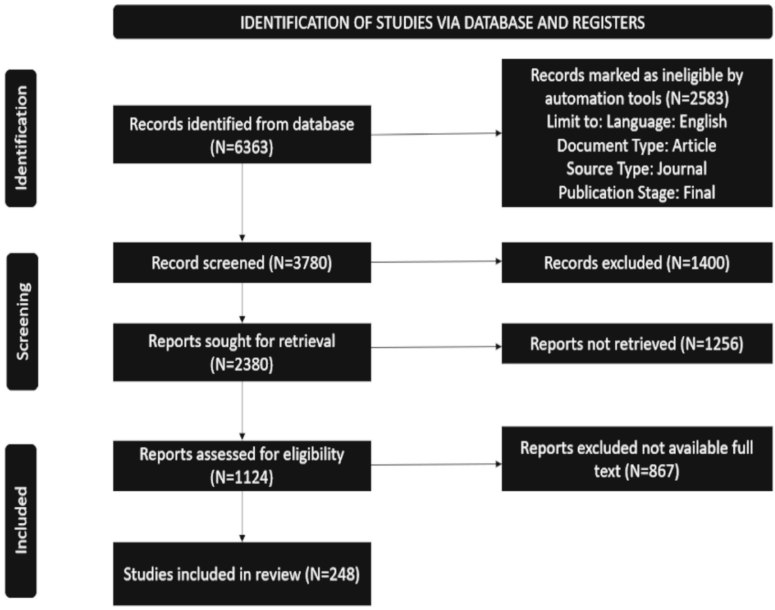


Fig. 2. Identification of relevant literature

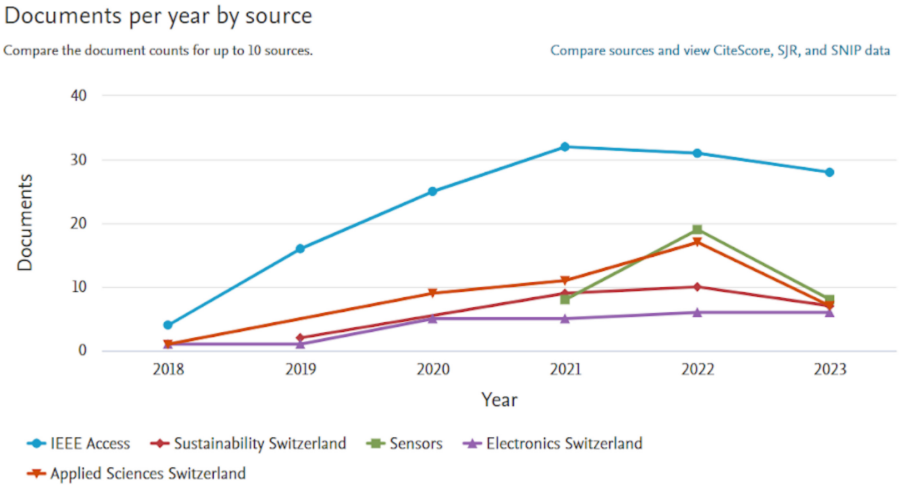


Fig. 3. Year wise Publication (Source: Scopus data-base)

Initial data was analysed with the help of Scopus Fig. 3 describes the number of articles published till July 2023 by varied sources; IEEE Access published highest number of papers in this area.

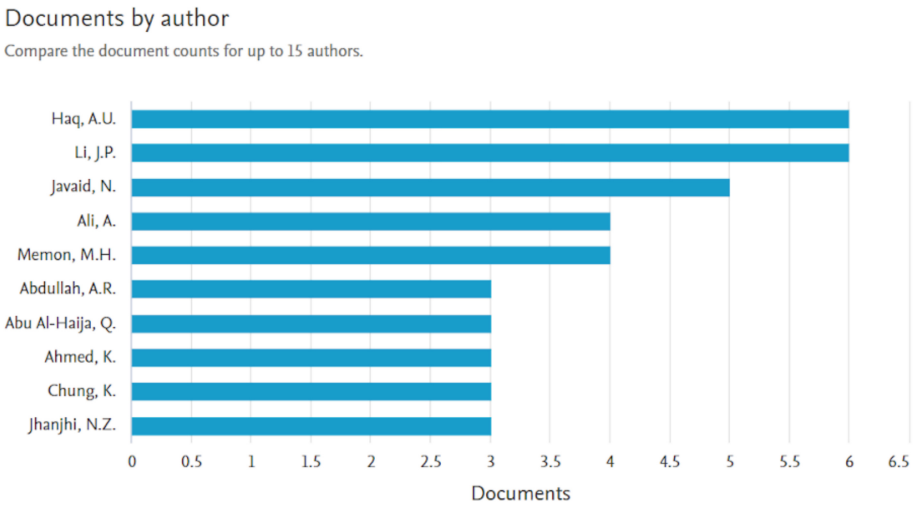


Fig. 4. Collaborative authors (Source: Scopus)

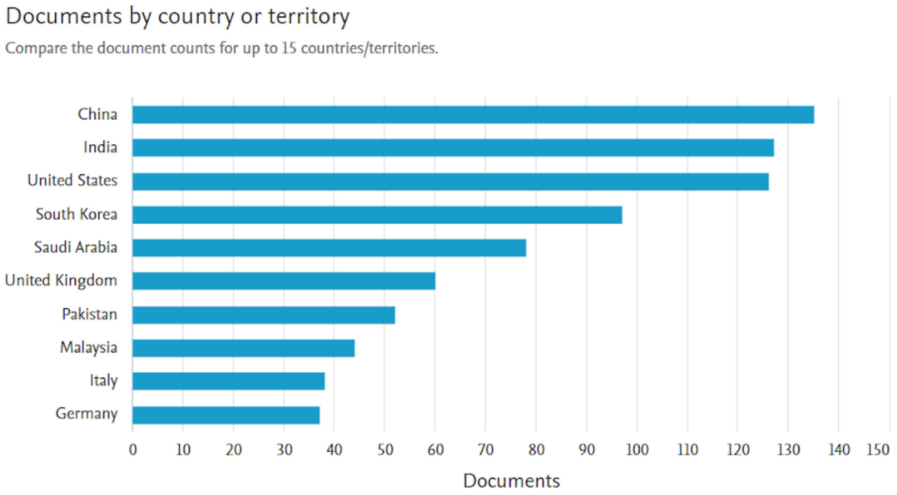


Fig. 5. Country wise publications (Source: Scopus)

The author-wise contribution (Fig. 4) shows that Haq, Amin Ul and Li, Jianping from China published 6 articles each depicted as most contributing author. In the same line Fig. 5 shows China as the most contributing country.

4 Systematic Literature Review Analysis with the Respect to Paper Publication Collaboration

As shown in Fig. 6 that papers review till 2023 authored as (36 as first author 14%, 38 as second author 15%,32 as third author 12% as follows) shows significant contribution in the form of paper publication in the field of Machine learning in with respect to the performance measurement matrix (Fig. 7).

Year/ Auth	1 Author	2 Author	3 Author	4 Author	5 Author	6 Author	7 Author	8 Author	Total
2016									0
2017			4						4
2018	2	5	3			6	7		23
2019	4		2					5	11
2020	5	5	4	4	7	8			33
2021	8	9	4	9	10	9	9	8	66
2022	9	10	8	7	9	7	6	4	60
2023	8	9	7	6	5	4	6	6	51
	36	**38**	**32**	**26**	**31**	**34**	**28**	**23**	**248**

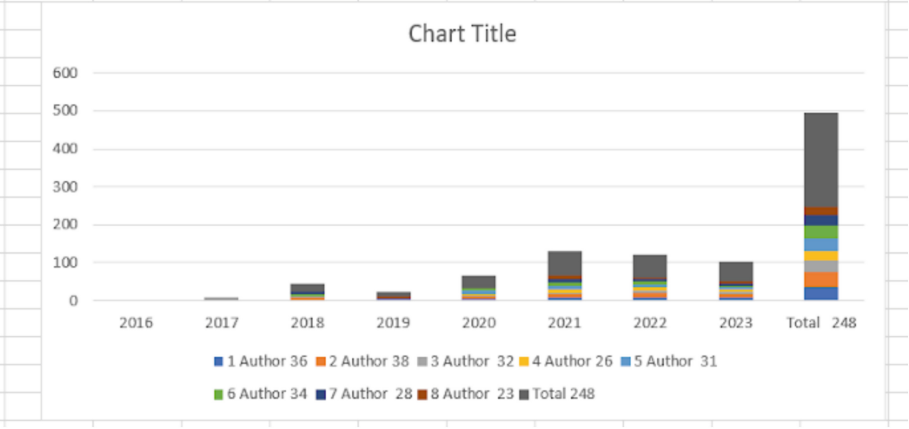


Fig. 6. Collaborative Authors (Source: Scopus)

As you can see from the figures above, we used Scopus database, to conduct our systematic literature review. We also made use of VOSviewers, a free piece of software that employs bibliometric networks to produce co-citation and co-authorship maps. In the references, this map shows which publications, authors, and research subjects are the most important network visualisation with co-occurrence (Avenali et al., 2023). In the field of performance measurement, the incorporation of machine learning has opened up new research avenues that can be explored with the help of VOSviewers. This led to the execution of a keyword analysis. There were a total of 112 keywords found. There can be no more than 32 occurrences of anything. Figure 6 is a diagram depicting the relationships between the most important keywords after selecting a minimum frequency of their

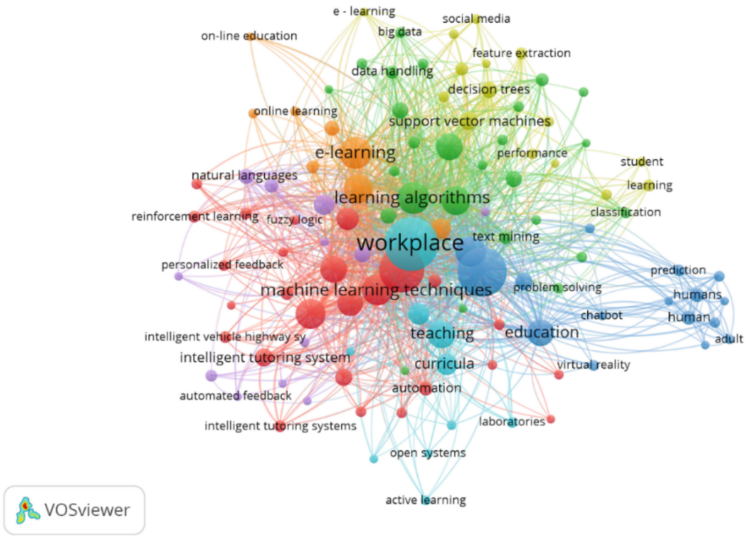


Fig. 7. Co-occurrence Analysis

occurrence. The significance and prevalence of a word or phrase informed the choice of colour to represent it. The blue cluster's most common key word is "workplace", where it has occurred 32 times with associated keywords. Performance measurement relies heavily on the contributions of the cluster's authors and co-authors; therefore, it makes sense that machine learning would be a powerful driver for the entire cluster as a whole. All clusters describe the relevance of literature in explaining role of machine learning on assessing employee's performance and also give directions for future research areas (Table 1).

Table 1. Cluster and Items and Cluster Theme

Cluster 1	Red	26 Items	Learning System, Machine Learning, Techniques, Deep Learning, Computer Aided Instructions, Personalised Feedback, Peer Review Grading, Automated Assessment, Intelligent Vehicle, Highway System	Learning Systems
Cluster 2	Green	26 Items	Learning Algorithm, Data Handling, Sentiment Analysis, Data Visualisation, Student Feedback, Classification of Information, Opinion Mining	Sentiment Analysis
Cluster 3	Blue	16 Items	Machine Learning, Education, Chatbot, Human, Prediction, Virtual Reality, Skill, Artificial Intelligence	Machine Learning
Cluster 4	Yellow	14 Items	Support Vector Machine, Decision Tree, Feature Extraction, Performance, Learning, Student, Social-Media, Random Forest	Performance
Cluster 5	Purple	12 Items	Natural Languages, Semantics, Student Response, Feedback to Workplace, Predictive Analysis, Modeling Languages	Predictive Analytics
Cluster 6	Sky Blue	10 Items	Work Place, Active Learning, Computer Software, Curricula, Teaching, Open System	Work Place
Cluster 7	Orange	8 Items	E-learning, Educational Data Mining, Educational Computing, online learning	e-learning

References

Avenali, A., Daraio, C., Di Leo, S., Matteucci, G., Nepomuceno, T.: Systematic reviews as a metaknowledge tool: caveats and a review of available options. Int. Trans. Oper. Res. **30**(6), 2761–2806 (2023). https://doi.org/10.1111/itor.13309

Broadus, R.N.: Toward a definition of "bibliometrics." Scientometrics **12**(5–6), 373–379 (1987). https://doi.org/10.1007/bf02016680

Choudhury, P., Allen, R.T., Endres, M.G.: Machine learning for pattern discovery in management research. Strateg. Manag. J. **42**(1), 30–57 (2020). https://doi.org/10.1002/smj.3215

Coates, J.B., Davis, E.W., Emmanuel, C.R., Longden, S.G., Stacey, R.J.: Multinational companies performance measurement systems: international perspectives. Manag. Account. Res. **3**(2), 133–150 (1992). https://doi.org/10.1016/s1044-5005(92)70008-0

Cravero, A., Pardo, S., Sepúlveda, S., Muñoz, L.: Challenges to use machine learning in agricultural big data: a systematic literature review. MDPI AG (2022). https://doi.org/10.20944/preprints202202.0345.v1

Fallucchi, F., Coladangelo, M., Giuliano, R., William De Luca, E.: Predicting employee attrition using machine learning techniques. Computers **9**(4), 86 (2020). https://doi.org/10.3390/computers9040086

Hawkins, M.: Virtual employee training and skill development, workplace technologies, and deep learning computer vision algorithms in the immersive metaverse environment. Psychosociol. Issues Hum. Resource Manag. **10**(1), 106 (2022). https://doi.org/10.22381/pihrm10120228

Hofmann, E., Brunner, J.H., Holschbach, E.: Research in business service purchasing: current status and directions for the future. Manag. Rev. Q. **70**(3), 421–460 (2019). https://doi.org/10.1007/s11301-019-00172-7

Kamble, S.S., Gunasekaran, A.: Big data-driven supply chain performance measurement system: a review and framework for implementation. Int. J. Prod. Res. **58**(1), 65–86 (2019). https://doi.org/10.1080/00207543.2019.1630770

Lyons, N.: Talent acquisition and management, immersive work environments, and machine vision algorithms in the virtual economy of the metaverse. Psychosociol. Issues Hum. Resource Manag. **10**(1), 121 (2022). https://doi.org/10.22381/pihrm10120229

Mody, M.A., Hanks, L., Cheng, M.: Sharing economy research in hospitality and tourism: a critical review using bibliometric analysis, content analysis and a quantitative systematic literature review. Int. J. Contemp. Hosp. Manag. **33**(5), 1711–1745 (2021). https://doi.org/10.1108/ijchm-12-2020-1457

Nadler, A., McGuigan, L.: An impulse to exploit: the behavioral turn in data-driven marketing. Crit. Stud. Media Commun. **35**(2), 151–165 (2017). https://doi.org/10.1080/15295036.2017.1387279

Nithya, B., Ilango, V.: Predictive analytics in health care using machine learning tools and techniques. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), June 2017. https://doi.org/10.1109/iccons.2017.8250771

Pehar, F.: From statistical bibliography to bibliometrics. Libellarium: Časopis Za Istraživanja u Području Informacijskih i Srodnih Znanosti **3**(1), 1–28 (2011). https://doi.org/10.15291/libellarium.v3i1.144

Sahinbas, K.: Employee promotion prediction by using machine learning algorithms for imbalanced dataset. In: 2022 2nd International Conference on Computing and Machine Intelligence (ICMI), 15 April 2022. https://doi.org/10.1109/icmi55296.2022.9873744

Sharma, A., Sharma, T.: HR analytics and performance appraisal system. Manag. Res. Rev. **40**(6), 684–697 (2017). https://doi.org/10.1108/mrr-04-2016-0084

Shilbayeh, S., Abonamah, A.: Predicting student enrolments and attrition patterns in higher educational institutions using machine learning. Int. Arab J. Inf. Technol. **18**(4) (2021). https://doi.org/10.34028/18/4/8

Wang, M., Cui, Y., Wang, X., Xiao, S., Jiang, J.: Machine learning for networking: workflow, advances and opportunities. IEEE Netw. **32**(2), 92–99 (2018). https://doi.org/10.1109/mnet.2017.1700200

Winarno, Widyatmojo, P.: The influences of human capital organizational learning and organizational competence on performance. In: Proceedings of the International Conference of Business, Economy, Entrepreneurship and Management (2019). https://doi.org/10.5220/0009967605670573

WebGPU: Comparing Parallelism Over Serial Execution in Web Graphics

M. Mallegowda[1(✉)], Tejas Hegde[1], Sini Anna Alex[2], and Anita Kanavalli[3]

[1] Department of CSE, M S Ramaiah Institute of Technology, Bangalore, India
mallegowdam@msrit.edu
[2] Department of CSE (AI and ML), M S Ramaiah Institute of Technology, Bangalore, India
sinialex@msrit.edu
[3] Department of AIDS, M S Ramaiah Institute of Technology, Bangalore, India
anithak@msrit.edu

Abstract. With ever-increasing use of the internet, web applications have become attractive. This is because web applications offer zero installation and a portable interface across several devices. From E-Commerce, banking systems and content management systems to video editing, file storage, remote terminals, video conferencing and scientific computations, webapps are more featureful than ever.

Traditional web apps could not support graphics intensive applications like 3D modelling, gaming and simulation. Such applications typically require an additional hardware device called GPU (Graphics Processing Unit) to accelerate graphics. WebGL, a web abstraction of the OpenGL library, was the first attempt to overcome this inability and enabled accelerated graphics on browsers. WebGPU, the successor of WebGL, provides additional features when compared to WebGL. In this work, WebGPU is described. The marching cubes algorithm is run using the CPU intensive Ring of Staging Buffer technique and then parallelised using the GPU accelerated Compute Shader method. Observations from these two techniques are compared. The results show that WebGPU provides significantly greater performance (nearly 6 times better) compared to serial execution.

Keywords: WebGPU · Ring-Buffer · Compute-Shader

1 Introduction

Webapps have become popular due to their simplicity. Users can easily access them through their web browsers. They need not spend time installing the software on their computer. The developers' work is also simplified. Webapps are platform agonistic and hence they save time on building system architecture/platform specific features for their software.

Graphics intensive applications like 3D modelling, gaming and simulation require high amounts of CPU and RAM. However, as time progressed, it was found that using CPU alone was insufficient and expensive to render high performance graphics. Hence the GPU (Graphics Processing Unit) was invented. It is a dedicated graphics rendering

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 219–224, 2026.
https://doi.org/10.1007/978-3-031-98138-8_18

unit. The excellent parallel processing ability of a GPU where a core may handle up to ten threads allows high performance graphics rendering.

Previously, to render GPU accelerated graphics on the web, developers had to develop plugins for the targeted browser, which could access the GPU. This was hardware dependent. To solve this problem, the OpenGL library was analysed. This software provided hardware independent access to the GPU, allowing cross platform applications which used GPU accelerated graphics to be developed. As a result, the WebGL API (application programming interface) was developed.

WebGPU is a successor of WebGL. It provides two crucial benefits over WebGL:

i. WebGPU is set to receive updates going forwards while WebGL will not.
ii. WebGPU can handle General Purpose GPU (GP-GPU) computations while WebGL cannot.

This means that WebGPU can not only be used for Graphics but for other applications such as Machine Learning.

2 Related Work

Introduction to WebGPU API and *Introduction to Computer Graphics and Ray-Tracing using the WebGPU API* courses by Benjamin Kenwright at SIGGRAPH 2022 are the only credible research materials found on the WebGPU API. However, no attempt was made to compare parallelism over serial execution in web graphics.

MARCHING CUBES: A HIGH RESOLUTION 3D SURFACE CONSTRUCTION ALGORITHM by Lorenson et al. at SIGGRAPH 1987, describes the Marching Cubes algorithm for surface reconstruction. In this work, we use the Marching Cubes algorithm to assess the performance of the CPU and the GPU based rendering methods.

Compute Shaders for Physics Processing by Mike Bailey at SIGGRAPH 2012 describes the GPU based Compute Shader execution technique which is the parallel execution-based rendering technique used in this work (Fig. 1).

3 Design

The following are the components of a general WebGPU model:

1. Physical GPU: which maybe integrated (built in), discrete(external) or software based. It is better to use a physical CPU over a software one.
2. A native GPU API sends instructions to the GPU and receives responses from it via the driver.
3. A WebGPU adapter is a virtual representation of the underlying GPU along with the native driver. To access the GPU, we need to create an adapter and then request access to the physical GPU via the native driver.
4. A logical device is an abstraction through which a browser can access the physical GPU in a compact way. Logical devices can be thought of as multiplexers which give webapps access to a physical device's GPU in isolation from other webapps for providing consistent, secure and logically correct results.

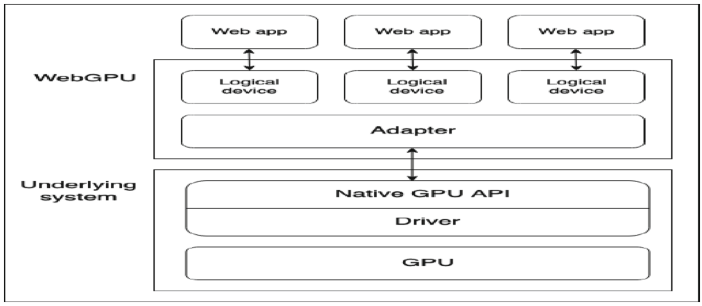


Fig. 1. General Model of WebGPU.

4 Experiment

The aim of the experiment is to prove the superiority of GPU based parallel methods over serial methods in web graphics. For this experiment, the Marching Cubes algorithm is used to perform surface triangulation of dynamic objects which are part of a continuous animation being produced on the browser. The performance gain obtained when this algorithm is ported onto a GPU is measured.

The metric used to assess the performance is the FPS (Frames Per Second). It is the rate at which consecutive frames are displayed i.e., the number of frames displayed per second. Higher the FPS, higher the performance of the graphics rendering method.

The browser accesses the on-device GPU through the logical device. There are three steps to access the GPU using JavaScript:

i. Use *navigator.gpu* to check if WebGPU is supported.
ii. Access an adapter via the *navigator.gpu.requestAdapter()* method.
iii. Request device via the *adapter.requestDevice()* method.

If the third step is successfully completed, the browser has access to the physical GPU.

Then, the CPU intensive method of Ring of Staging Buffers is applied on the marching cubes algorithm. In this method, data is read/written to the GPU memory. Staging Buffers before the read/write operation takes place. Since buffers are written to very frequently, a ring of staging buffers is used. The FPS is measured for this method.

After that, the GPU accelerated Compute Shader method is applied on the marching cubes algorithm. This method eliminates the need for staging buffer by direct generation of data in the GPU. The data is divided into smaller chunks and all these chunks are parallelly processed by a large number of GPU cores. The FPS is measured for this method.

The difference in the FPS between the two methods is then calculated. This gives the performance boost that WebGPU offers over CPU execution.

5 Experimental Results and Observations

The experiment was performed on an ASUS TUF F-15 Gaming Laptop with integrated Nvidia GeForce GTX 1650 GPU having dedicated 4 GB memory.

The Brave browser (Chromium based) with Chromium Version 113 was used on the Windows platform (Fig. 2).

5.1 Rendering Method: Ring of Staging Buffers

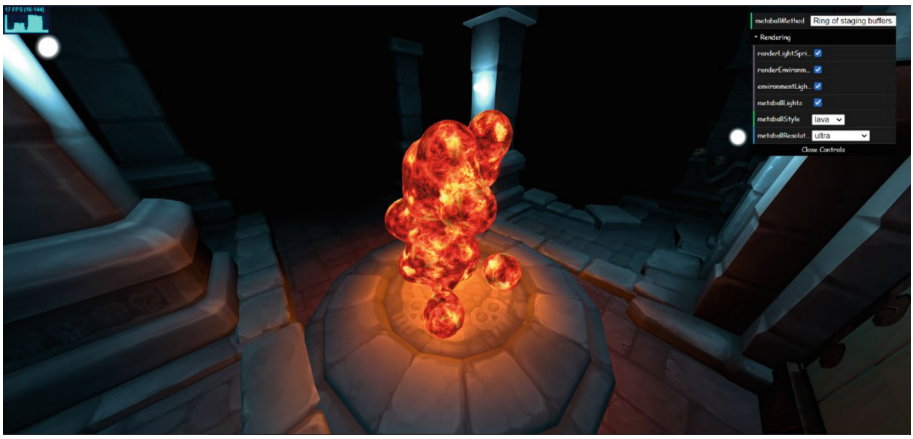


Fig. 2. CPU Intensive Rendering Method

This rendering method produces an *average FPS of 17 fps* (Fig. 3).

5.2 Rendering Method: Compute Shader

Fig. 3. GPU accelerated rendering method

This rendering method produces an *average FPS of 100 fps*. During the experiment, it was observed that the CPU intensive method causes high resolution graphics to be

sluggish. This fact is reinforced by the low FPS rate of 17 fps for CPU intensive method. However, at the same resolution, the GPU accelerated method provides a very smooth graphics rendering which is evident by the very high FPS rate of 100 fps. The difference in the FPS rates is $100 - 17 = 83\,fps$. Hence the performance increase in percentage due to parallelism over serial execution is:

$$\frac{\text{Difference in FPS}}{(\text{CPU intensive FPS})} * 100 = \frac{83}{17} * 100 = 588.24\%$$

Hence WebGPU provides nearly 6 times greater performance compared to the CPU based methods. This shows how parallelism gives superior performance over serial execution.

6 Compatibility

WebGPU was first introduced in Google Chrome 113. As of now, only the latest versions of Google Chrome and Microsoft Edge support the WebGPU API fully. Also, it is only supported in the following operating systems: ChromeOS, macOS, and Windows. WebGPU may have issues with low GPU RAM. However, as computers get more powerful and cheaper, with ever increasing usage of the internet, the WebGPU API will only get better in terms of its functionality, browser compatibility and OS supportability.

7 Conclusion

WebGPU is an API which allows modern browsers to access physical GPUs and accelerate graphics on the web. GPU accelerated graphics use parallel computation methods which provide a huge performance increase over serial execution. It can not only be used for graphics; it can also be used in Machine Learning applications for boosting the prediction/inference process. WebGPU has the potential to pull graphics intensive applications like 3D modeling, simulation and gaming from the Desktop to the web. Hence WebGPU is one of the foundations of the future web.

References

1. Kenwright, B.: Introduction to WebGPU API. In: SIGGRAPH Conference 2022
2. Kenwright, B.: Introduction to computer graphics and ray-tracing using the WebGPU API. In: SIGGRAPH Conference (2022)
3. Lorensen, W.E., Cline, H.E., Cubes, M.: A high resolution 3D surface construction algorithm. In: SIGGRAPH Conference (1987)
4. Bailey, M.: Compute shaders for physics processing. In: SIGGRAPH Conference (2012)
5. Va, H., Choi, M.-H., Hong, M.: Real-time cloth simulation using compute shader in unity 3D for AR/VR contents. MDPI Appl. Sci. (2021)
6. Schütz, M., Kerbl, B., Wimmer, M.: Rendering point clouds with compute shaders and vertex order optimization. Wiley Comput. Graph. Forum (2021)
7. Mallet, I., Yuksei, C.: Deferred adaptive compute shading. In: HPG 2018: Proceedings of the Conference on High-Performance Graphics (2018)

8. Dharma, D., Jonathan, C., Kistidjantoro, A.I., Manaf, A.: Material point method based fluid simulation on GPU using compute shader. In: 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), Denpasar, Indonesia, pp. 1–6 (2017). https://doi.org/10.1109/ICAICTA.2017.8090962
9. Tornai, R., Fürjes-Benke, P.: Compute shader in image processing development. In: Proceedings of the 1st Conference on Information Technology and Data Science, Debrecen, Hungary, 6–8 November 2020
10. Bilodeau, B.: Efficient compute shader programming. In: Game Developers Conference, San Francisco, CA (2011)
11. Junker, A., Palamas, G.: Real-time interactive snow simulation using compute shaders in digital environments. In: FDG 2020: Proceedings of the 15th International Conference on the Foundations of Digital Games (2020)
12. Schutz, M., Wimmer, M.: Rendering point clouds with compute shaders. In: SA 2019: SIGGRAPH Asia (2019)
13. OpenGL. https://en.wikipedia.org/wiki/OpenGL
14. WebGL. https://en.wikipedia.org/wiki/WebGL
15. WebGPU API Documentation. https://developer.mozilla.org/en-US/docs/Web/API/WebGPU_API
16. WebGPU Chrome. https://developer.chrome.com/blog/webgpu-io2023/
17. Compute Shader WebGPU Chrome. https://developer.chrome.com/articles/gpu-compute/
18. WebGPU Metaballs Demo. https://toji.github.io/webgpu-metaballs/
19. Ring of Staging Buffers and Compute Shader. https://toji.dev/webgpu-best-practices/buffer-uploads
20. Compute Shader. https://media.siggraph.org/education/conference/S2012_Materials/ComputeShader_1pp.pdf

Stock Market Prediction Technique Through LSTM and NLP

Shirish Joshi[1(✉)], Harsh Chauhan[1], and Sahil Kulkarni[2]

[1] Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India
shirish.joshi@sicsr.ac.in
[2] Deloitte Deutschland, Potsdam, Germany

Abstract. The basic idea behind the paper is to develop a method to analyze how the people receive the changes in the stock market and how it affects the stock's closing price in the future. The stock is a fraction of ownership that a person can purchase and benefit from the company's growth in the market. There are many factors involved in the change of stock price, one of the major factors is the company's next step in the market and how it is received by the people. The stock market data values vary from time to time according to the subject risk. Therefore, it is the need to develop a computational automated methodology to predict the stock market data values. The major goals of the proposed initiative are to assist investors, encourage safe stock market investing, and enhance those goals. In addition to this, the system has an algorithm that can assess patterns in stock prices.

Keywords: Stock Market · Prediction Technique · Machine Learning · Support Vector Machine · Long Short Term Memory

1 Introduction

In today's world, investment has become a buzzword. Everyone who is earning even a small amount is thinking of investing the money to secure their future financial requirements and to take care of the retired life. When it comes to investment, the first thing that comes to mind is the investment in the stock market. So far, many people have invested in the stock market and gained lot of profit. On the other side, many people have lost money in the stock market. This is because they invest without studying anything about the stock market at all. Before investing in the stock market, one should study the market very well and then invest. There are many factors involved in the change of stock price, one of the major factors is the company's next step in the market and how the people receive it. The stock price of a company can be predicted if one studies the movement in the market using models, which are correctly designed, developed, and refined. Due to dynamic nature, predicting the exact stock price is very difficult. The major goals of the proposed initiative are to assist investors, encourage safe stock market investing, and enhance those goals. In addition to this, the system has an algorithm that can assess

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 225–237, 2026.
https://doi.org/10.1007/978-3-031-98138-8_19

patterns in stock prices. (HAYES, 2022) [3]. Stock market prediction systems have long been an essential tool for stock traders [14]. Stock market is the most popular investment scheme which yields high returns but with some risks [16, 19]. Stock trading is dynamic and volatile in nature which makes the task of stock market trend prediction a complex problem [17].

Existing Technique

A. The Average Directional Index (ADX): An example of a trend indicator, the ADX is used to gauge the trend momentum of a company. Trading in the trend's direction boosts returns while lowering risk.
B. The A/D line - This is one of the most widely used indicators that uses supply and demand information to forecast whether or not people are buying the stock. (MITCHELL) [5].

1.1 Limitation of Existing Technique

All the techniques, which exist in the real world, are plotting of previous data and a simple regression technique to predict further. There is no such technique that can analyze peoples' points of view on overstock.

Existing websites

1. MoneyControl (http://www.moneycontrol.com/) [23] - this website has basic stock analysis live market data and updating news.
2. TickerTape (https://www.tickertape.in/) [24] - this website helps the user to find historical data and also gives a stock forecast.
3. Investing.com (https://in.investing.com/) [25]-this website is a bit different from the previous one as it uses a feature called stock screener helps the user find the market cap of the company.

1.2 Limitation of Existing Websites

The existing websites are using basic analysis for prediction on the contrary proposed project uses advanced Machine Learning techniques like LSTM [1, 12, 15, 18] and NLP [20, 22] for prediction and analysis of the stock, which provides a more practical, and trustworthy prediction base for the stock.

2 Literature Review

In the realm of finance, trading stocks is a crucial activity. Studies have used the Support Vector Machine (SVM) technique of machine learning [7, 8] to forecast stock prices for large and small financial organizations, as well as in three different markets, employing prices with daily and minutely frequency (Reddy, 2018).

According to the weak version of the Efficient Market Hypothesis (EMH) [2], it is challenging to forecast an asset's future price using previous prices. Because of this, the market behaves erratically, making predicting difficult. Financial forecasting is a difficult task because of the financial system's intrinsic complexity.

The objective was to use artificial intelligence (AI) [4] techniques to model and forecast the price of an index on the stock market in the future. Three artificial intelligence techniques are used to forecast the price of a stock market index based on historical price data: neural networks (NN), support vector machines (SVM), and neuro-fuzzy systems. Because they can account for the complexity of the financial system, artificial intelligence algorithms are used as financial time series forecasting tools. This makes use of the random walk (RW), the linear modeling approach, and the autoregressive moving average (ARMA). Using information from the Johannesburg Stock Exchange, the study was continued. A series of previous All Share Index closing prices served as the basis for the data.

In order to demonstrate the Internet of Multimedia of Things for stock analysis, the project's goal was to develop a novel neural network approach. The data was obtained from the livestock market for real-time and offline analysis as well as from the results of historical visualizations and analyses. While researching the impact of market factors on stock prices, traditional neural network algorithms may incorrectly predict the stock market because the initial weight of the random selection problem is easily susceptible to inaccurate predictions.

3 Problem Formulation

3.1 The Objective of the Proposed System

A. To make a prediction on whether or not a stock will increase - The major goal of using LSTM [9, 21] and NLP analysis is to anticipate whether the stock will increase and try to pinpoint by how many points it will climb or decline. The major goal is to present a novel deep neural network-based method that can assist investors in predicting the ups and downs of the company.
B. To make stock market investments safer - To make the stock market safer so that new investors may make decisions based on trustworthy analysis and profit from them.
C. The project's major goal is to develop a new stock research technique that has never been utilized and that complies with contemporary technology trends, automating a portion of the work of a stockbroker and an investor.

3.2 Feasibility Study

A. Technological Feasibility - The project is a fully functioning recurrent neural network model, the following technologies will be used to establish the system

- Datalore - powerful Jupiter base idle for data science and Machine Learning algorithm construction
- Tiingo - A API used widely for collecting data for various stocks to date
- Spreadsheet - to manually clear and clean data for analysis. Tiingo and its APIs are built to be performant, and consistent and support extensive filters to speed up development time [11].

B. Resource and time availability

The proposed system needs the following resources

- Coding system - portable or regular desktop
- Programming tools - all are freely available except Datalore has a limitation on the number of notebooks created in one project
- Internet – a fast connection is needed for the quick functioning of Datalore

Time allocation for the project

- Planning and design - 2 days
- Designing the algorithm - 3 weeks
- Coding the proposed algorithms - 4 weeks
- Documentation - 3 days

C. Economic feasibility - The economic cost is very low or none, the main cost is the time, effort, and skill invested by me to develop the technique.
D. Possibility from a sociological and lawful standpoint - This technique being open source offers an insight into how a deep neural network can be included in the stock market, this will also help hundreds of brokers and investors in the market to invest their money safely with minimal risk.

4 Proposed Methodology

The basic idea as discussed before is to predict by how many points the stock price will rise or go down and use the impact of the company's liking in the public to do the same. We propose two methods to do so for predicting the points by how much the stock will go down or up:

4.1 Predictive Price Analysis

Since the model has four layers and can interact with each other, we will use recurrent neural networks to be more precise and to implement Long Short-Term Memory analysis. This will allow us to store all of the model's previously learned crucial information.

The main concept is to anticipate the stock point and to use the data far back from Covid 19, so that machine can also learn the cycle of the stock reaction. To reach the result, we use the data dated back to 17 April 2017. Tiingo is used to import the data prior to Covid so that the model can get updates automatically.

To expedite the process, we will attempt to import the model from TensorFlow and Keras rather than creating the model manually. Additionally, these models have been evaluated, are error-free, and are highly accurate. We suggest splitting the data set in half in the ratio of 60:40 in order to estimate the model's future stock point after 28 days. Here are some of the steps involved in determining the stock point after 28 days:

A. Data Import and cleaning - To import the data as discussed before we will use Tiingo API so that the CSV can be updated in real-time, this also reduces the burden of data cleansing because the data cleaning process is already performed by API, with only some minor slicing and indexing of the data will be needed.

B. Data Preprocessing - For data preprocessing, we import sklearn's min-max scaler, the library helps us scale down the values of the dataset in the range between 0 and 1. This speeds up the process of Machine Learning and helps the model to bring the result in a small period.

   Then the scaled data is converted into an array using the library NumPy to further facilitate the model implementation.

C. Model Implementation - For model implementation, we import the model from TensorFlow directly we implement the data directly into the function and the desired result is achieved.

D. Model plotting and presentation - To see the model's result, we will use Matplotlib to plot the result and import this model.

4.2 Predictive Sentiment Analysis

For finding the people's opinion of the company, we analyze the company's news which will help us identify how the company's next move in the market is received by the people. (Ohashi, 25) We propose the following NLP analysis technique.

A. Data Import and Cleaning - We import data from Kaggle. Unfortunately, there is no way to update this data set in real-time as there is no API still available or any other way to do this. The only alternative is to construct a completely new API that specifically points out a company's news. The data is cleaned manually using Excel all the duplicate news, blank spaces are removed from the dataset.

B. Data Preprocessing - We remove the punctuation from the headline then, we convert the words into lowercase so that the model accuracy can be increased. After that, we divide headlines into tokens so that we can put the words into an array, then we remove the stop-words from the headline.

C. Sentiment analysis - Then we allot a polarity score and a subjective score. To make it a comprehensive score, we use both of them and combine them so that if the sentiment score is less than 0, we consider it a negative review,

   If the sentiment score is equal to 0, it is a neutral emotion [6] and, if the sentiment score is more than 0, it is a positive review we count all the reviews.

D. Model Plotting - To see the result, we plot the final count of each company on a bar graph to instantly understand the company's standing.

The basic idea is to get the results of both analyses and combine them into a UI that will tell the closing point of the stock and also tell whether the stock will rise or not based on the company's news analysis.

5 FAANG Stock Prediction Technique – (Facebook, Amazon, Apple, Netflix, Google)

The following screenshots show the LSTM (Long Short Term Memory) analysis and NLP (Natural Language Processing). The image is updated after the closing time (Fig. 1).

5.1 Facebook

Long Short Term Memory (LSTM) Analysis of Facebook
A plot of the closing prices for Facebook shares as anticipated by the Long Short Term Memory analysis model can be seen in the first image from the left. The model's database is represented by the blue line, its training data is represented by the yellow line, and its forecast for the following 28 days is shown by the green light. On a scale of 0 to 1, the expected closing price is 1.2498997 higher. Taking into account the natural language processing analysis for the company's news, the final analysis predicts that the stock will rise for the Facebook Corporation.

Natural Language Processing (NLP) Analysis
The graph displays the results of the company's news' Natural Language Processing analysis, which aids in determining whether the market will make a good or negative turn in the future. According to the graph, the news about the firms has largely been neutral, and the positive news about the companies outweighs the negative news. As a result, the study comes to the conclusion that the company's next action will raise the value of its shares (Fig. 2).

5.2 Amazon

Long Short Term Memory (LSTM) Analysis of Amazon
A plot of the closing prices for Amazon stock as anticipated by the Long Short Term Memory Analysis model can be seen in the first image from the left. The model's database is represented by the blue line, its training data is represented by the yellow line, and its forecast for the following 28 days is shown by the green light. On a scale of 0 to 1, the expected closing price is 2.91 up. In light of the natural language processing study for the company's news, the final analysis determines that the stock will rise for the Facebook Corporation.

Fig. 1. Current image showing Long Short Term Analysis (LSTM) and Natural Language Processing (NLP) analysis of Facebook. (adapted from attached file Facebook.pdf)

Fig. 2. Current image showing Long Short Term Analysis (LSTM) and Natural Language Processing (NLP) analysis of Amazon. (adapted from attached file Amazon.pdf)

Natural Language Processing (NLP) Analysis

The analysis of the company's news using natural language processing is shown in the above graph, which aids in determining whether the market will make a good or negative turn in the future. According to the graph, the news about the firms has largely been neutral, and the positive news about the companies outweighs the negative news. As a result, the study comes to the conclusion that the company's next action will raise the value of its shares (Fig. 3).

5.3 Apple

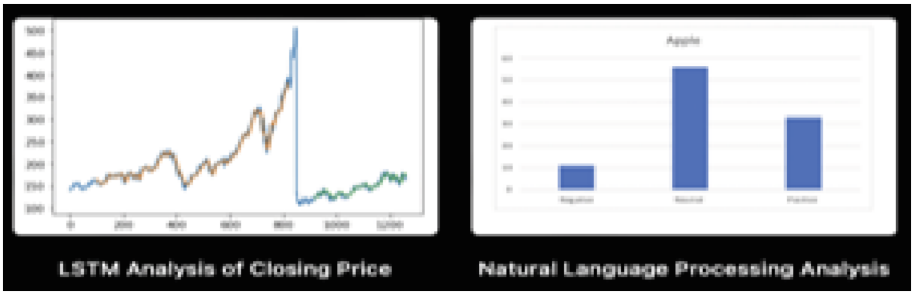


Fig. 3. Current image showing Long Short Term Analysis (LSTM) and Natural Language Processing (NLP) analysis of Apple. (adapted from attached file Apple.pdf)

Long Short Term Memory (LSTM) Analysis of Apple

A plot of the closing prices for Apple stock as anticipated by the Long Short Term Memory Analysis model can be seen in the first image from the left. The model's database is represented by the blue line, its training data is represented by the yellow line, and its forecast for the following 28 days is shown by the green light. On a scale

of 0 to 1, the expected closing price is 0.2279 higher. In light of the natural language processing study for the company's news, the final analysis determines that the stock will rise for the Facebook Corporation.

Natural Language Processing (NLP) Analysis
The analysis of the company's news using natural language processing is shown in the above graph, which aids in determining whether the market will make a good or negative turn in the future. According to the graph, the news about the firms has largely been neutral, and the positive news about the companies outweighs the negative news. As a result, the study comes to the conclusion that the company's next action will raise the value of its shares (Fig. 4).

5.4 Netflix

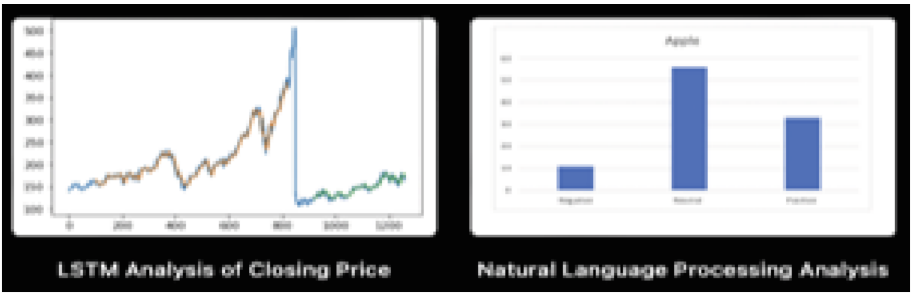


Fig. 4. Current website showing Long Short Term Analysis (LSTM) and Natural Language Processing (NLP) analysis of Netflix. (adapted from attached file Netflix.pdf)

Long Short Term Memory (LSTM) Analysis of Netflix
A plot of the closing prices for Netflix stock as anticipated by the Long Short Term Memory Analysis model can be seen in the first image from the left. The model's database is represented by the blue line, its training data is represented by the yellow line, and its forecast for the following 28 days is shown by the green light. On a scale of 0 to 1, the expected closing price is 0.28136 higher. In light of the natural language processing study for the company's news, the final analysis determines that the stock will rise for the Facebook Corporation.

Natural Language Processing (NLP) Analysis
The analysis of the company's news using natural language processing is shown in the above graph, which aids in determining whether the market will make a good or negative turn in the future. According to the graph, the news about the firms has largely been neutral, and the positive news about the companies outweighs the negative news. As a result, the study comes to the conclusion that the company's next action will raise the value of its shares (Fig. 5).

5.5 Google

Fig. 5. Current website showing Long Short Term Analysis (LSTM) and Natural Language Processing (NLP) analysis of Google.

Long Short Term Memory (LSTM) Analysis of Netflix

A plot of the closing prices for Google shares as anticipated by the Long Short Term Memory Analysis model can be seen in the first image from the left. The model's database is represented by the blue line, its training data is represented by the yellow line, and its forecast for the following 28 days is shown by the green light. On a scale of 0 to 1, the expected closing price is 4.560 up. In light of the natural language processing study for the company's news, the final analysis determines that the stock will rise for the Facebook Corporation.

Natural Language Processing (NLP) Analysis

The analysis of the company's news using natural language processing is shown in the above graph, which aids in determining whether the market will make a good or negative turn in the future. According to the graph, the news about the firms has largely been neutral, and the positive news about the companies outweighs the negative news. As a result, the study comes to the conclusion that the company's next action will raise the value of its shares.

6 Results

The study concludes that the stock price is most likely going to increase if the firm has more good headlines than the negative headlines and the LSTM analysis reveals that the company's stock is on a rising path. The FANG firms that we chose show that their stock values are constantly growing.

Due to the company's recent decision to control its loss during Covid 19, each of the company's stock closing prices after 28 days show an increase. Additionally, this enables us to confirm that the analysis is accurate.

To go in detail further.

Facebook: The stock price of Facebook will grow by 1.24 points, and sentiment research reveals that the news about the social network has mostly been viewed as neutral and that more people have given it favorable than negative feedback.

Amazon: The stock price of Amazon will grow by 2.91 points, while sentiment research reveals that Facebook news has gotten a majority of neutral reactions and that good reviews outnumber negative ones, indicating that the stock price of Facebook will undoubtedly increase.

Apple: The price of Apple's stock will grow by 0.22 points, and sentiment research reveals that Facebook's announcement was mostly viewed as neutral and that more people gave it favorable than negative feedback, indicating that Facebook's stock price would undoubtedly increase.

Netflix - Netflix's stock price will grow by 0.28 points, and sentiment research reveals that Facebook's announcement was mostly seen as neutral, with more positive than negative comments. These findings indicate that Facebook's stock price will undoubtedly increase.

Table 1. Test Cases showing expected and actual result

Step #	Step Details	Expected Result	Actual Result	Pass/Fail/Not Executed/Suspended
1	Predicting the closing point	Site should show closing point of predicted stock after 3 am every day	As Expected	Pass
2	Predicting the Company's likability	Site should show how companies news have been received by the public over a span of a month	As Expected	Pass

Google - The stock price of Google will grow by 4.560 points, and sentiment research reveals that the news about Facebook has been mostly regarded as neutral and that good reviews have outweighed negative ones, indicating that the stock price of Facebook will undoubtedly increase (Table 1).

7 Test Cases

The aim which we wanted to achieve was predicting the closing point of the stock, the site should update the prediction every day at 3 am The site achieved this as expected, the second task is to predict the company's likability among the people, the site showed the company's likability status depending on the company's news.

8 Future Scope of Development

Future applications for this kind of methodology include not just the sentiment analysis of headlines but also the sentiment analysis of social media opinions about the firm, which might improve forecasting accuracy.

We may also create a live news updating platform where the news dataset is updated in real-time, allowing us to include the most recent events in the analysis. Additionally, it will offer a more accurate prediction method for current market events. This will help in getting the more accurate information about the increase or decrease in the stock price and hence give investors a better opportunity for investment and earn higher profits.

For the prediction, we used only one correct variable, but in the future, we can also use different variables like a Twitter reaction towards the company or some other way through which we can capture a more wide-ranging perspective of the company's sentiments.

9 Conclusion

The project demonstrates that using the right amount of data and the right variables can help us predict a number, whose value is dependent on many variables. The sentiment analysis of the company can help us in prediction of the stock prices in a much better way and with more accuracy. The negative, neutral and positive headlines about any company can be processed using the live data which will further help to accurately guess the company's growth and hence the share prices of that company. This will help the investors to create more investment opportunities and invest more and more and earn high returns. Long Short Term Memory (LSTM) and Natural Language Processing (NLP) techniques can be effectively used to process the data effectively and come up with better conclusions about the share prices.

References

1. Colah: Understanding-LSTMs. Retrieved from Colah blogs, 27 August 2015. https://en.wikipedia.org/wiki/Long_short-term_memory
2. Downey, L.: Efficient Market Hypothesis (EMH), 31 December 2021. www.invetopedia.com. https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp
3. Hayes, A.: Stock Definition (2022). https://www.investopedia.com/. https://www.investopedia.com/terms/s/stock.asp
4. Marwala, L.R.: Forecasting the stock market index using artificial intelligence technique. In: Forecasting the Stock Market Index Using Artificial Intelligence Technique, vol. 116 (n.d.)

5. Mitchell, C.: Advance/Decline (A/D) Line. Retrieved from Advance/Decline (A/D) Line, 1 January 2022. https://www.investopedia.com/terms/a/advancedeclineline.asp
6. Ohashi, R.M.: From sentiment analysis to emotion recognition: a NLP story, 25 July 2015. https://medium.com/. https://medium.com/neuronio/from-sentiment-analysis-to-emotion-recognition-a-nlp-story-bcc9d6ff61ae
7. Reddy, V.K.: Stock Market Prediction Using Machine Learning, vol. 5 (2018)
8. Mathanprasad, L., Gunasekaran, M.: Analysing the trend of stock market and evaluate the performance of market prediction using machine learning approach. In: 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, pp. 1–9 (2022)
9. Sisodia, P.S., Gupta, A., Kumar, Y., Ameta, G.K.: Stock market analysis and prediction for Nifty50 using LSTM deep learning approach. In: 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 156–161 (2022)
10. Wang, Y., Wang, Y.: Using social media mining technology to assist in price prediction of stock market. In: 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, pp. 1–4 (2016)
11. Periketi, S., Kulkarni, R., Dulhare, U.N.: LSTM and prophet model fusion framework for bitcoin candlestick visualization and price prediction forecasting. In: 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 1190–1194 (2023)
12. Malim, T.N.A.B.T., Kamarudin, S.A., Ahad, N.A., Mamat, N.A.M.G.: Prediction of FTSE bursa Malaysia KLCI stock market using LSTM recurrent neural network. In: 2022 IEEE International Conference on Computing (ICOCO), Kota Kinabalu, Malaysia, pp. 415–418 (2022)
13. Sharma, A., Bhuriya, D., Singh, U.: Survey of stock market prediction using machine learning approach. In: 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, pp. 506–509 (2017)
14. Sakphoowadon, S., Wisitpongphan, N., Haruechaiyasak, C.: Probabilistic lexicon-based approach for stock market prediction: a case study of the stock exchange of Thailand (SET). In: 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, Thailand, pp. 383–388 (2018)
15. Li, H., Cao, Y., Yang, X., Wang, Y.: Model optimization for stock market prediction using multiple labelling techniques. In: 2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, pp. 161–165 (2022)
16. Somani, P., Talele, S., Sawant, S.: Stock market prediction using hidden Markov model. In: 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference, Chongqing, China, pp. 89–92 (2014)
17. Jindal, R., Bansal, N., Chawla, N., Singhal, S.: Improving traditional stock market prediction algorithms using Covid-19 analysis. In: 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, pp. 374–379 (2021)
18. Gunturu, P.A., Joseph, R., Revant, E.S., Khapre, S.: Survey of stock market price prediction trends using machine learning techniques. In: 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, pp. 1–5 (2023)
19. Iyer, M., Mehra, R.: A survey on stock market prediction. In: 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan, India, pp. 663–668 (2018)
20. Sandhya, P., Bandi, R., Himabindu, D.D.: Stock price prediction using recurrent neural network and LSTM. In: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 1723–1728 (2022)

21. Jeenanunta, C., Chaysiri, R., Thong, L.: Stock price prediction with long short-term memory recurrent neural network. In: 2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), Khon Kaen, Thailand, pp. 1–7 (2018)
22. Veena, R.C., Shivakanth, G.: Stock price prediction using LSTM and TLBO. In: 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, pp. 1–5 (2023)
23. http://www.moneycontrol.com/
24. https://www.tickertape.in/
25. https://in.investing.com/

Author Index

J. R. Saini et al. (Eds.): ICCAML 2024, CCIS 2538, pp. 239–240, 2026.
https://doi.org/10.1007/978-3-031-98138-8