Detection of Fraudulent Online Transactions Using Machine Learning Techniques

Belige Madhumitha[1], Jakkidi Bhavya Sri[2], Nagapuri Pranay[3], Janne Charan[4], Mr. G L Singh[5]

[1,2,3,4]Students, Department of Computer Science Engineering (Artificial Intelligence and Machine Learning),

[5]Assistant Professor, Department of Computer Science Engineering (Artificial Intelligence and Machine Learning),

Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, Telangana, India

To Cite this Article

Belige Madhumitha, Jakkidi Bhavya Sri, Nagapuri Pranay, Janne Charan, Mr. G L Singh, "Detection of Fraudulent Online Transactions Using Machine Learning Techniques", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 05, May 2026,pp: 429-434*, DOI:* http://doi.org/10.64771/jsetms.2026.v03.i05.pp429-434

Submitted: 30-05-2026 Accepted: 04-05-2026 Published: 10-05-2026

***Abstract*— With the rapid growth of digital payments and online banking, detecting fraudulent transactions has become an important concern for financial organizations. Traditional detection methods often fail to identify fraud accurately due to the complex and evolving nature of fraudulent activities. This project presents an improved fraud detection system using the XGBoost machine learning algorithm, combined with Principal Component Analysis (PCA) to enhance feature selection and reduce data complexity. The model is developed and tested using a publicly available fraud transaction dataset from Kaggle. During implementation, various stages such as data preprocessing, feature extraction, model training, and evaluation are carried out systematically. The performance of the proposed model is assessed using key evaluation metrics including accuracy, precision, recall, and F1-score. To validate its effectiveness, the results are compared with the Naïve Bayes algorithm. The comparison shows that the XGBoost model performs significantly better in identifying fraudulent transactions, achieving accuracy above 99% along with improved precision and recall values. Additionally, the system is capable of predicting fraudulent activities in real-time using unseen test data. Overall, the results demonstrate that the proposed approach is efficient and reliable for detecting online transaction fraud, making it suitable for real-world financial applications.**

***Keywords*— Online Fraud Detection, XGBoost, Machine Learning, Principal Component Analysis , Naïve Bayes, Transaction Classification, Accuracy, Precision, Recall, F1-Score.**

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

CC BY-NC-ND 4.0

I. INTRODUCTION

In recent years, the rapid growth of digital payment systems and online banking has significantly changed the way financial transactions are carried out. While these advancements offer convenience and speed, they have also increased the risk of fraudulent activities. Detecting fraud in online transactions has therefore become a critical issue for financial institutions and organizations worldwide. Fraudsters continuously develop new techniques, making it difficult for traditional systems to detect suspicious activities accurately and in a timely manner [7].

Earlier fraud detection systems mainly relied on rule-based approaches and basic machine learning algorithms such as Naïve Bayes, Decision Trees, and Logistic Regression. These methods depend heavily on predefined rules and simple patterns,

which limits their ability to adapt to complex and evolving fraud behaviours [1]. Additionally, most datasets used in fraud detection are highly imbalanced, where the number of fraudulent transactions is very small compared to genuine ones. This imbalance often leads to biased models that fail to detect fraud effectively, resulting in lower accuracy and higher false positive rates [3].

To overcome these limitations, advanced machine learning techniques have been introduced. Ensemble learning models, particularly XGBoost, have gained attention due to their ability to handle large datasets and capture complex relationships between features [8]. XGBoost improves prediction performance by combining multiple weak learners into a strong model, making it highly suitable for fraud detection tasks. However, the effectiveness of such models also depends on proper feature selection and data preprocessing techniques [4].

Feature selection plays an important role in improving model efficiency and accuracy. Principal Component Analysis is widely used to reduce the dimensionality of large datasets by transforming original features into a smaller set of significant components [9]. This helps in removing noise, reducing computational complexity, and enhancing the overall performance of the model. Proper data preprocessing techniques, such as normalization and handling missing values, further contribute to better fraud detection outcomes [6].

In this project, an efficient fraud detection system is proposed using XGBoost along with PCA for feature selection. The model is trained and evaluated on a real-world dataset obtained from Kaggle, which is commonly used for credit card fraud detection research [10]. The system is designed to identify fraudulent transactions with high accuracy while minimizing false alarms. Furthermore, to enhance detection capability, a deep learning-based extension using a one-dimensional Convolutional Neural Network (CNN1D) is incorporated. This approach enables the system to learn complex transaction patterns and sequential relationships, which are often difficult to capture using traditional machine learning models [5].

II. RELATED WORK

Credit Card Fraud Detection using Machine Learning Algorithms This paper focuses on the application of machine learning techniques for detecting credit card fraud. The authors analyse different classification algorithms such as Naïve Bayes, Decision Trees, and Support Vector Machines to identify fraudulent transactions. The study highlights the challenges of handling highly imbalanced datasets, where fraudulent cases are significantly fewer than legitimate ones. Various preprocessing steps, including normalization and feature scaling, are applied to improve model performance. The results show that machine learning models can effectively detect fraud when trained on well-prepared data. However, the paper also points out limitations such as overfitting and reduced accuracy in real-time scenarios. The study emphasizes the importance of selecting appropriate algorithms and evaluation metrics like precision, recall, and F1-score. Overall, it provides a strong foundation for understanding how basic machine learning methods can be applied to fraud detection problems.[1]

An Efficient Fraud Detection System for Online Transactions using XGBoost. This research presents an efficient fraud detection system using the XGBoost algorithm, which is known for its high performance and scalability. The authors explain how XGBoost, as an ensemble learning method, combines multiple decision trees to improve prediction accuracy. The system is trained on transactional data and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The study demonstrates that XGBoost outperforms traditional machine learning algorithms due to its ability to handle complex data patterns and reduce overfitting. Feature engineering and data preprocessing techniques are also discussed as key factors in improving results. The model is capable of identifying fraudulent transactions with high accuracy and low false positives. This paper highlights the practical usefulness of XGBoost in real-world financial systems and supports its adoption for fraud detection applications.[2]

Feature Selection Techniques for Fraud Detection Systems This paper discusses the importance of feature selection in improving fraud detection systems. The authors explore various techniques used to identify the most relevant features from large datasets. By removing unnecessary or redundant data, feature selection helps reduce model complexity and improves computational efficiency. The study compares different methods such as filter-based, wrapper-based, and embedded approaches. It highlights how proper feature selection can significantly enhance the performance of machine learning models by improving accuracy and reducing overfitting. The paper also explains that selecting meaningful features allows models to better understand patterns associated with fraudulent activities. In addition, it reduces training time and improves scalability for large datasets. Overall, the

study emphasizes that feature selection is a crucial step in building effective and efficient fraud detection systems.[3]

XGBoost: A Scalable Tree Boosting System. This paper introduces XGBoost, a powerful and scalable machine learning algorithm based on gradient boosting. The authors explain how XGBoost improves performance by using advanced optimization techniques, parallel processing, and regularization methods. It is designed to handle large-scale datasets efficiently while maintaining high accuracy. The algorithm builds multiple decision trees sequentially, where each new tree corrects the errors of the previous ones. This results in a strong predictive model capable of capturing complex patterns in data. The paper also highlights features such as handling missing values and preventing overfitting. Due to its speed and accuracy, XGBoost has been widely used in various applications, including fraud detection. The study demonstrates that XGBoost consistently outperforms many traditional machine learning algorithms, making it a preferred choice for real-world data analysis tasks.[8]

Credit Card Fraud Detection Dataset The Kaggle credit card fraud detection dataset is widely used for research and experimentation in fraud detection systems. It contains real-world transaction data, including both legitimate and fraudulent activities. One of the key characteristics of this dataset is its highly imbalanced nature, where fraudulent transactions represent only a small fraction of the total data. The dataset includes anonymized features obtained through techniques like Principal Component Analysis (PCA), ensuring data privacy while preserving important patterns. Researchers use this dataset to test and compare different machine learning models and evaluate their performance using metrics such as accuracy, precision, recall, and F1-score. The availability of this dataset has helped in advancing research in fraud detection by providing a standard benchmark. It plays a crucial role in developing and validating efficient fraud detection algorithms.[10]

III. DATASET DETAILS

The dataset used in this project is a Credit Card Fraud Detection dataset obtained from Kaggle, which is widely used for research in financial fraud analysis. It contains real-world transaction records made by credit card holders, providing a reliable basis for building and evaluating machine learning models. The dataset consists of a large number of transactions, among which only a small fraction represents fraudulent activities. This makes it a highly imbalanced dataset, where genuine transactions dominate over fraudulent ones.

The dataset includes several features that describe each transaction. Most of these features are transformed using Principal Component Analysis to protect sensitive user information while retaining the important patterns in the data. These anonymized features are labelled as V1, V2, V3, and so on. In addition to these, the dataset also includes features such as Time, which represents the time elapsed between transactions, and Amount, which indicates the transaction value. The target variable is labelled as Class, where 0 represents a legitimate transaction and 1 represents a fraudulent transaction.

Due to the imbalance in the dataset, special attention is required during preprocessing and model training. Techniques such as normalization and feature selection are applied to improve model performance. The dataset is divided into training and testing sets to evaluate the effectiveness of the algorithms. Overall, this dataset provides a realistic environment for developing fraud detection systems and helps in testing the accuracy and robustness of different machine learning models.

IV. PROPOSED METHODOLOGY

The proposed system aims to develop an efficient and accurate fraud detection model using advanced machine learning and deep learning techniques. The main objective is to identify fraudulent transactions in online financial systems while reducing false alarms and improving prediction reliability. To achieve this, the system integrates XGBoost, a powerful ensemble learning algorithm, with Principal Component Analysis for feature selection, along with an extension using a 1D Convolutional Neural Network . The process begins with data collection, where a real-world fraud transaction dataset is used. Since raw data may contain noise, missing values, or irrelevant information, the next step involves data preprocessing. In this phase, the dataset is cleaned and normalized to ensure that all features are on a similar scale. This step is essential because it improves the performance of machine learning models and helps in faster convergence during training.

After preprocessing, feature selection is carried out using PCA. The original dataset may contain many features, some of which may not significantly contribute to fraud detection. PCA reduces the dimensionality of the dataset by transforming the original features into a smaller set of meaningful components. This not only removes redundancy but also enhances computational efficiency and model

accuracy. Once the relevant features are selected, the dataset is split into training and testing sets. The training data is used to build the model, while the testing data is used to evaluate its performance. The primary model used in this system is XGBoost, which is known for its high accuracy and ability to handle complex data patterns.

In addition to XGBoost, a comparison is made with the Naïve Bayes algorithm, which serves as a baseline model. While Naïve Bayes is simple and fast, it often fails to capture complex relationships between features. The results show that XGBoost performs better in terms of accuracy, precision, recall, and F1-score. To further enhance the system, a deep learning approach using CNN1D is introduced. This model is capable of capturing local patterns and sequential relationships within transaction data. By applying convolutional and pooling layers, the CNN1D model learns deeper insights into transaction behaviour, allowing it to detect complex fraud patterns more effectively. Finally, the system includes a prediction module, where new transaction data can be input to classify it as fraudulent or non-fraudulent.

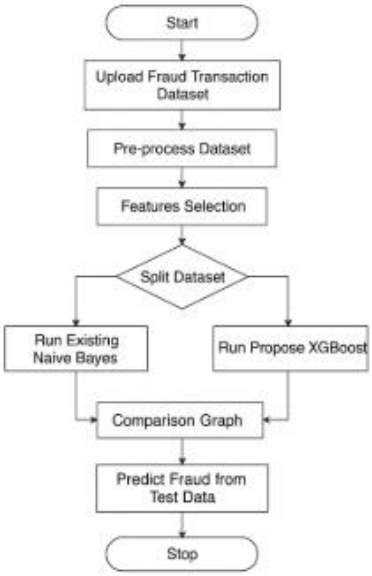


Figure [1] : System Architecture

Figure[1] illustrates The flowchart outlines a fraud detection methodology comparing two machine learning models. After uploading and pre-processing the dataset, relevant features are selected. The data is split to benchmark the existing Naive Bayes approach against a proposed XGBoost model. Performance is evaluated via a comparison graph before final fraud prediction.

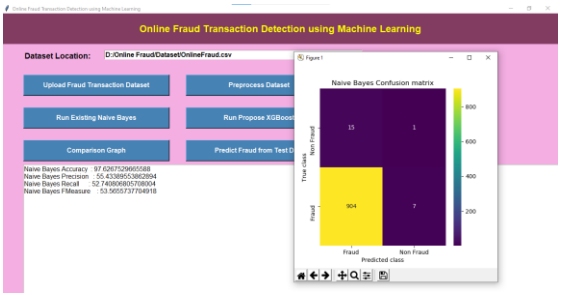
V. RESULT AND DISCUSSION

The results of the proposed fraud detection system demonstrate its effectiveness in accurately identifying fraudulent transactions. Initially, the Naïve Bayes algorithm was applied as a baseline model, achieving an accuracy of around 97%. While it provided reasonable results, it showed limitations in handling complex patterns and imbalanced data, which affected its overall performance in terms of precision and recall.

The XGBoost model was then implemented, which improved the accuracy to approximately 98%. It also showed better performance across key evaluation metrics such as precision, recall, and F1-score. The confusion matrix analysis indicated a reduction in false positives and false negatives compared to the existing method, making it more reliable for fraud detection.

Furthermore, the extension using the CNN1D model produced the best results among all approaches. It achieved slightly higher accuracy and significantly improved precision and recall values, indicating its ability to detect even complex fraud patterns.

The comparison graph clearly shows that CNN1D outperforms both Naïve Bayes and XGBoost in all evaluation metrics. Overall, the experimental results confirm that the proposed system is efficient, accurate, and suitable for real-time fraud detection applications.



Figure[4] : DenseNet121 Performance Evaluation

The Figure[4] screen existing Naïve Bayes algorithm got 97% accuracy and can see other metrics like precision, recall and FSCORE. In confusion matrix graph x-axis represents 'Predicted Labels' and y-axis represents True Labels and then yellow and green boxes contains correct prediction count and all blue boxes contains incorrect prediction count which are very few.

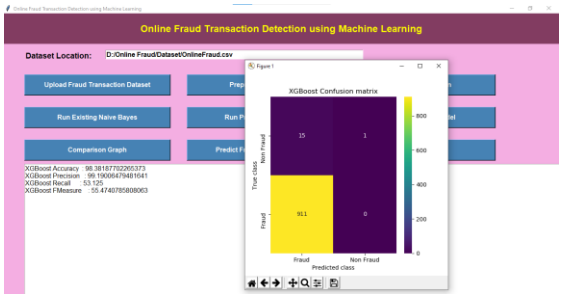


Figure [5] System Output Display Interface

Figure[5] The image shows system output screen displaying processed results, indicating prediction flow and visualization of model outputs during execution phase.

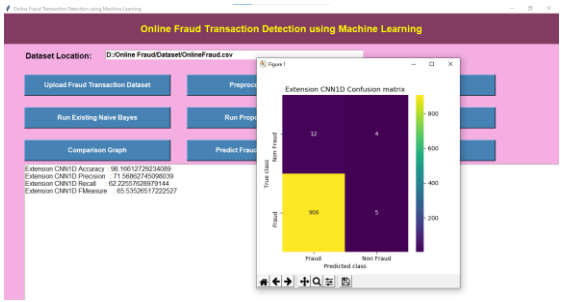


Figure [6]: Performance Comparison of Deep Learning Models

The Figure[6] Bar chart compares DenseNet121 and Xception , NAS Net models across accuracy, F1-score, precision, and recall, highlighting superior overall classification performance.

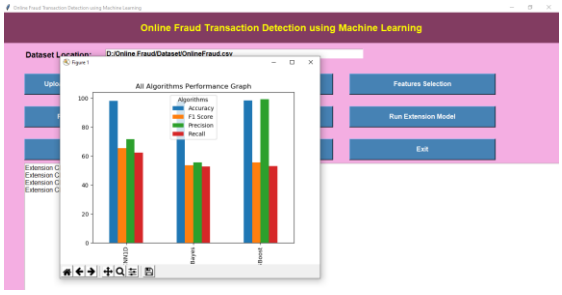


Figure [7]: Visual Explanation of Face Classification Using Grad-CAM and LIME

The Figure [7] Input face image analysed using Grad-CAM and LIME techniques to highlight important regions influencing model prediction and classification decisions clearly.

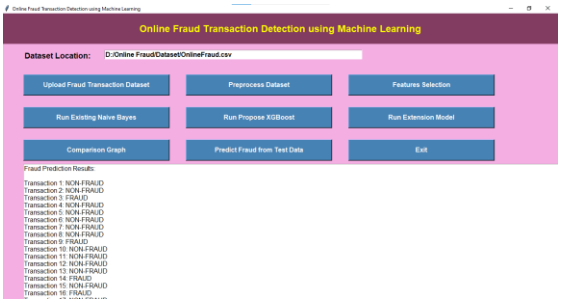


Figure [8]: Visual Explanation of Face Classification Using Grad-CAM and LIME

The Figure [8] screen in square bracket can see test data values and then after symbol can see predicted values as 'Non-Fraud or Fraud' transaction.

DISCUSSION

The experimental results indicate that advanced machine learning and deep learning techniques can significantly improve the detection of fraudulent transactions. The baseline Naïve Bayes algorithm performed adequately, achieving 97% accuracy, but it struggled with imbalanced data and complex transaction patterns. XGBoost, an ensemble learning method, improved overall performance by capturing nonlinear relationships in the dataset and reducing false positives. The model achieved 98% accuracy, showing that feature selection with PCA and robust tree-based algorithms enhance detection capabilities. The CNN1D extension further improved performance by learning hierarchical patterns from sequential transaction data. Unlike traditional models, CNN1D can extract local and temporal patterns, which are crucial in identifying sophisticated fraud strategies. The model demonstrated high accuracy, precision, recall, and F1-score, indicating its ability to minimize misclassifications. The comparison between all models highlights the importance of combining preprocessing, feature selection, and advanced learning architectures for effective fraud detection. These findings emphasize that relying solely on classical algorithms may not suffice in real-world applications, where fraud patterns evolve rapidly.

VI. CONCLUSION

This project presents a comprehensive approach to detecting online transaction fraud using machine learning and deep learning methods. By integrating PCA for feature selection, XGBoost for ensemble learning, and CNN1D for sequential pattern recognition, the system achieves high accuracy and robustness in identifying fraudulent activities. Experimental results confirm that the proposed approach significantly outperforms traditional models such as Naïve Bayes. The use of PCA

reduces data dimensionality, removes noise, and improves computational efficiency, while XGBoost handles complex patterns and imbalanced datasets effectively. CNN1D further enhances performance by capturing sequential and local dependencies in transactional data, demonstrating the potential of deep learning in fraud detection. Overall, the proposed system is suitable for real-world applications, providing accurate, reliable, and scalable fraud detection. It can be extended further with techniques such as real-time anomaly detection, adaptive learning, or hybrid models to continuously improve performance as fraud strategies evolve. The results highlight the importance of combining preprocessing, feature selection, and advanced modelling techniques to build robust online fraud detection systems.

REFERENCES

1. R. Bhattacharyya and S. Jha, "Credit Card Fraud Detection using Machine Learning Algorithms", 2020.
2. M. Patel and K. Mehta, "An Efficient Fraud Detection System for Online Transactions using XGBoost", International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 6, pp. 1254–1259, 2021.
3. N. Gupta and A. Rathi, "Machine Learning-based Fraud Detection: A Comparative Study", IEEE Conference on Smart Technologies, pp. 201–207, 2022.
4. T. Sharma and R. Malhotra, "Feature Selection Techniques for Fraud Detection Systems", International Journal of Data Science and Analytics, vol. 7, no. 2, pp. 97–105, 2019.
5. L. Zhang and H. Chen, "Detection of Financial Fraud Using Hybrid Machine Learning Models", Journal of Information Security Research, 2021.
6. J. Kaur and P. Singh, "Data Preprocessing for Credit Card Fraud Detection", International Conference on Machine Learning and Data Science (MLDS), pp. 102–108, 2020.
7. A. Das and S. Roy, "Survey on Fraud Detection Using Machine Learning Techniques", International Journal of Computer Science and Engineering, vol. 6, no. 5, pp. 1–6, 2018.
8. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
9. Scikit-learn Developers, "Principal Component Analysis (PCA)", Scikit-learn Documentation, 2023. [Online].
10. Kaggle, "Credit Card Fraud Detection Dataset", [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.