AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 3(2026)

Hybrid Fraud Detection in Digital Transactions Using XGBoost and 1D Convolutional Neural Networks

JONNADA BHAVANI GOUD[1], Dr.S.SWATHI RAO[2]
[1]Student, Department of Computer Science and Engineering, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.
[2]Assistant Professor, Department of AI & ML, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.
[2]Email: swathirao1511@gmail.com

Abstract— **The increased popularity of digital payment systems has resulted in efficient and convenient online financial operations. On the other hand, such development requires more efficient fraud detection methods as more opportunities have appeared for fraudsters to conduct their illegal operations. This project suggests utilizing a hybrid method of fraud detection based on the use of PCA, XGBoost and CNN1D. PCA helps to remove unimportant features and improve the quality of input data. XGBoost algorithm is designed to detect complicated dependencies between the transaction features, while CNN1D is used to learn hidden features which can be used to distinguish real and fraudulent transactions. The experiment was conducted using the publicly available financial transaction dataset divided into train and test data samples in 70:30 proportion. Accuracy, precision, recall, F1-score and confusion matrix were used to evaluate the performance of the model. The obtained experimental results prove that the suggested hybrid model is capable of detecting fraud accurately and reliably.**

Keywords—**Online Fraud Detection, Financial Transaction Security, XGBoost, CNN1D, Principal Component Analysis (PCA), Machine Learning, Deep Learning, Fraud Classification, Digital Payment Systems, Transaction Risk Analysis.**

I. INTRODUCTION

The rapid expansion of digital technologies has transformed the way financial transactions are carried out across the world. Online banking, mobile wallets, internet shopping, and digital payment gateways have become an essential part of everyday life due to their speed and convenience. Individuals and businesses increasingly depend on these platforms to transfer money, purchase goods, and access financial services from any location. While these advancements have improved accessibility and reduced transaction time, they have also introduced new security challenges. Cybercriminals continuously develop sophisticated methods to exploit vulnerabilities in digital payment systems, resulting in unauthorized transactions and financial losses. Traditional security measures, including passwords and rule-based verification systems, are often unable to detect evolving fraud techniques. As transaction volumes continue to increase, financial institutions require intelligent and automated approaches that can identify suspicious activities quickly and accurately while minimizing the impact on legitimate customers **[1], [7]**.

Fraud detection has become one of the most significant concerns for financial organizations because fraudulent transactions can lead to severe economic losses, reputational damage, and reduced customer confidence. Modern fraud schemes often involve identity theft, account takeover, phishing attacks, stolen payment credentials, and unauthorized fund transfers. These fraudulent activities frequently resemble legitimate transactions, making them difficult to identify using conventional methods. Manual verification is impractical due to the enormous number of transactions processed every second. Consequently, organizations are increasingly adopting data-driven techniques capable of analyzing transaction characteristics and recognizing hidden behavioral patterns. Machine learning provides an effective solution by learning from historical transaction records and automatically distinguishing normal activities from suspicious ones. These intelligent systems continuously improve their predictive ability as additional data become available, making them suitable for dynamic financial environments where fraud strategies constantly evolve **[1], [3], [7]**.

Machine learning algorithms have demonstrated considerable success in solving classification problems involving large and complex datasets. Among these techniques, XGBoost has gained recognition for its high predictive performance, scalability, and ability to manage structured financial data efficiently. It combines multiple

decision trees to create a powerful ensemble model capable of identifying subtle relationships between transaction attributes. Before training the model, data preprocessing and feature selection play an important role in improving performance. Principal Component Analysis (PCA) is widely used to eliminate redundant information and reduce the dimensionality of datasets while preserving essential characteristics. This process helps reduce computational complexity and minimizes the influence of irrelevant features. By combining effective preprocessing techniques with a robust classification algorithm, financial fraud detection systems can achieve improved accuracy, reduced false alarms, and faster decision-making during transaction monitoring **[2], [4], [8], [9]**.

Although machine learning methods perform well on structured financial datasets, deep learning techniques have shown remarkable capability in extracting complex patterns that may not be easily recognized by conventional algorithms. One-Dimensional Convolutional Neural Networks (CNN1D) are particularly effective for analyzing sequential and numerical data because they automatically learn hierarchical feature representations without requiring extensive manual feature engineering. In fraud detection, CNN1D captures subtle relationships among transaction attributes, enabling the identification of fraudulent behavior even when attackers modify their strategies. Integrating CNN1D with traditional machine learning techniques creates a hybrid framework that combines the strengths of both approaches. Such a framework benefits from the interpretability and efficiency of ensemble learning while leveraging the pattern recognition capabilities of deep learning. This combination enhances overall detection performance and provides greater resilience against emerging fraud techniques **[3], [5], [8]**.

This project proposes a hybrid fraud detection framework that integrates Principal Component Analysis (PCA), XGBoost, and CNN1D to improve the identification of fraudulent online financial transactions. The dataset undergoes preprocessing to handle missing values, normalize feature values, and prepare the data for effective model training. PCA is applied to select the most informative features, reducing unnecessary complexity and improving computational efficiency. The processed data are divided into training and testing sets using a 70:30 ratio. The performance of the proposed approach is evaluated using standard metrics, including accuracy, precision, recall, F1-score, and confusion matrix, and compared with a Naïve Bayes baseline model. The experimental results demonstrate that the hybrid framework effectively distinguishes legitimate and fraudulent transactions, providing a reliable, scalable, and practical solution for enhancing security in modern digital payment systems **[2], [5], [6], [9], [10]**.

II. RELATED WORK

"Credit Card Fraud Detection using Machine Learning Algorithms," R. Bhattacharyya and S. Jha [1]

This study explored the application of machine learning algorithms for detecting fraudulent credit card transactions. The authors compared multiple classification techniques to identify the most effective model for distinguishing legitimate and fraudulent activities. The research emphasized the importance of preprocessing financial transaction data and handling class imbalance to improve prediction accuracy. Experimental results demonstrated that machine learning models can significantly reduce fraudulent transactions while maintaining a low false positive rate. The study concluded that intelligent fraud detection systems can support financial institutions in strengthening transaction security and minimizing financial losses.

"An Efficient Fraud Detection System for Online Transactions using XGBoost," M. Patel and K. Mehta [2]

This research proposed an online fraud detection framework based on the XGBoost algorithm to improve the accuracy of fraud classification. The authors focused on utilizing ensemble learning techniques to analyze transaction patterns and identify suspicious activities in real time. The proposed approach demonstrated better prediction performance than traditional machine learning algorithms by effectively handling complex datasets. The study also highlighted the importance of feature engineering and data preprocessing in enhancing model efficiency. The findings suggested that XGBoost provides a reliable solution for developing secure and scalable fraud detection systems.

"Feature Selection Techniques for Fraud Detection Systems," T. Sharma and R. Malhotra [4]

This study investigated various feature selection methods used to improve the performance of fraud detection models. The authors examined how eliminating irrelevant and redundant features reduces computational complexity while increasing classification accuracy. Different dimensionality reduction techniques were evaluated to identify the most informative transaction attributes. The research demonstrated that appropriate feature selection

significantly enhances the efficiency of machine learning algorithms and minimizes overfitting. The study concluded that optimized feature selection plays a crucial role in developing accurate and computationally efficient fraud detection systems.

"Detection of Financial Fraud Using Hybrid Machine Learning Models," L. Zhang and H. Chen [5]

This research presented a hybrid machine learning framework for detecting financial fraud by combining multiple predictive algorithms. The authors aimed to improve fraud detection performance by integrating the strengths of different learning models instead of relying on a single classifier. The proposed hybrid approach successfully captured complex fraud patterns and achieved higher accuracy, precision, and recall than individual models. The study emphasized that combining multiple techniques improves the robustness of fraud detection systems and supports financial organizations in identifying sophisticated fraudulent activities more effectively.

"XGBoost: A Scalable Tree Boosting System," T.Chen and C. Guestrin [8]

This study introduced the XGBoost algorithm as a scalable and efficient gradient boosting framework designed for high-performance machine learning applications. The authors presented several optimization techniques, including parallel processing, regularization, and efficient tree construction, to improve prediction accuracy and computational speed. Extensive experiments demonstrated that XGBoost consistently outperformed many existing machine learning algorithms across various classification and regression tasks. The research established XGBoost as a reliable algorithm for handling large datasets and solving complex predictive problems, making it highly suitable for fraud detection and other real-world data analysis applications.

III. DATASET DETAILS

The dataset used in this project is the Credit Card Fraud Detection Dataset obtained from the Kaggle repository. It contains real-world financial transaction records collected for research on fraud detection and machine learning applications. The dataset consists of transactions performed by credit card users over a specific period and includes both legitimate and fraudulent transactions. Each transaction is represented by multiple attributes that describe different characteristics of the payment process. To protect user privacy and maintain confidentiality, most of the original transaction features have been transformed using Principal Component Analysis (PCA), resulting in anonymized variables labeled from V1 to V28. In addition to these transformed features, the dataset contains the Time attribute, which indicates the elapsed time between transactions, the Amount attribute representing the transaction value, and the Class attribute that serves as the target label. The Class value of 0 represents a legitimate transaction, while 1 indicates a fraudulent transaction. The dataset contains 284,807 transaction records with 31 columns, including 30 input features and one target class. Since fraudulent transactions form only a very small percentage of the total data, the dataset is highly imbalanced, making fraud detection a challenging classification problem.

Before training the machine learning and deep learning models, the dataset undergoes several preprocessing steps to improve its quality and ensure reliable prediction results. Initially, the dataset is examined for missing or inconsistent values, and appropriate preprocessing techniques are applied to prepare clean input data. Numerical features are normalized to reduce variations in feature scales and improve model convergence during training. Since most transaction attributes are already transformed using PCA, the dataset preserves important statistical information while maintaining customer anonymity. Additional feature selection is performed to identify the most informative attributes that contribute significantly to fraud detection, thereby reducing computational complexity and improving classification performance. After preprocessing, the dataset is randomly divided into 70% training data and 30% testing data. The training set is used to build the predictive models, while the testing set evaluates their ability to classify unseen transactions accurately. The performance of the proposed hybrid framework is measured using evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix**.** This publicly available dataset provides a reliable benchmark for comparing different fraud detection algorithms and enables researchers to develop intelligent models capable of identifying fraudulent financial transactions with high accuracy and reduced false prediction rates.

IV. PROPOSED METHODOLOGY

The proposed system introduces a hybrid fraud detection framework that combines data preprocessing, feature selection, machine learning, and deep learning techniques to improve the identification of fraudulent online financial

transactions. Initially, the uploaded transaction dataset is examined to remove inconsistencies and prepare it for model training. Missing values are handled appropriately, categorical information is converted into numerical format where required, and all feature values are normalized to maintain a consistent scale. After preprocessing, **Principal Component Analysis (PCA)** is applied to identify the most relevant features while removing redundant information that does not significantly contribute to fraud prediction. This process reduces the dimensionality of the dataset, lowers computational complexity, and improves model efficiency. The selected features are then divided into training and testing datasets using a **70:30 ratio**. The training dataset is used to build predictive models, while the testing dataset is reserved for evaluating their performance. Initially, the XGBoost algorithm is trained to learn complex relationships among transaction attributes through an ensemble of decision trees. This enables the system to recognize suspicious transaction patterns efficiently and produce accurate predictions for unseen financial transactions.

To further improve fraud detection capability, the proposed framework integrates a **One-Dimensional Convolutional Neural Network (CNN1D)** after feature selection. Unlike conventional machine learning models, CNN1D automatically learns hierarchical representations from transaction data by applying multiple convolution and pooling operations that capture hidden relationships among numerical features. This enables the model to identify subtle fraud patterns that may not be detected using traditional classification techniques alone. The trained CNN1D model is evaluated on the testing dataset, and its performance is compared with Naïve Bayes and XGBoost using standard evaluation metrics such as **accuracy, precision, recall, F1-score, and confusion matrix**. A comparison graph is generated to visualize the performance of all models and demonstrate the effectiveness of the proposed approach. Finally, the developed system provides a prediction module that accepts new transaction records and classifies them as either **fraudulent** or **legitimate** in real time. By combining efficient feature selection with the predictive strength of XGBoost and the pattern-learning capability of CNN1D, the proposed system offers a reliable, scalable, and accurate solution for enhancing the security of modern digital payment and online banking environments.

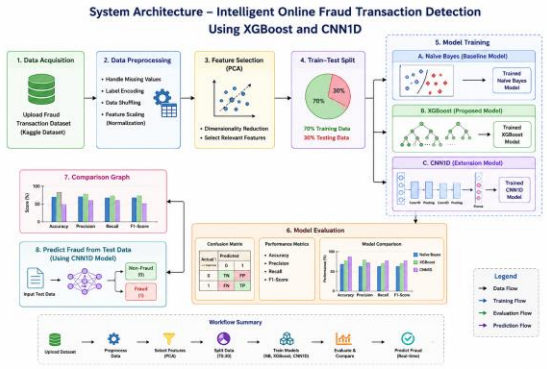

Figure 1: Proposed System Architecture for Hybrid Online Fraud Transaction Detection

Figure 1 illustrates the overall architecture of the proposed hybrid online fraud transaction detection system. The process begins with uploading the fraud transaction dataset collected from the Kaggle repository. The uploaded data undergoes preprocessing, where missing values are handled, categorical attributes are encoded, the dataset is shuffled, and feature values are normalized to improve data quality. Next, Principal Component Analysis (PCA) is applied to reduce dimensionality by selecting the most relevant features. The processed dataset is then divided into 70% training data and 30% testing data**.** During model training, three algorithms—Naïve Bayes (baseline model), XGBoost (proposed model), and CNN1D (extension model)—are trained and evaluated. Their performances are measured using Accuracy**,** Precision**,** Recall**,** F1-Score**,** and the Confusion Matrix, and the results are compared through a graphical analysis. Finally, the trained CNN1D model is used to classify new transaction records as either fraudulent or legitimate, providing an accurate and reliable solution for securing modern online financial transactions.

V.RESULT AND DISCUSSION

The proposed hybrid fraud detection system was successfully implemented and evaluated using the Kaggle Credit Card Fraud Detection dataset. After preprocessing and feature selection using Principal Component Analysis (PCA), the dataset was divided into 70% training data and 30% testing data. The performance of the proposed approach was compared with the existing Naïve Bayes classifier using evaluation metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. Experimental results showed that the XGBoost model achieved higher prediction performance than the Naïve Bayes algorithm, while the CNN1D extension model further improved the overall detection capability by learning complex transaction

patterns. The comparison graph demonstrated that CNN1D obtained the highest values across all evaluation metrics with fewer false predictions. The final prediction module accurately classified unseen transactions as either legitimate or fraudulent, confirming the effectiveness of the proposed hybrid framework. These results indicate that integrating PCA, XGBoost, and CNN1D provides a reliable, efficient, and scalable solution for detecting fraudulent online financial transactions with high accuracy.

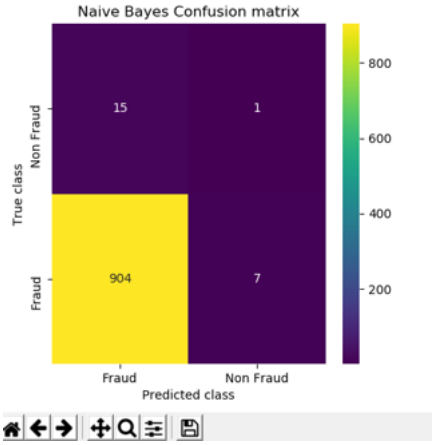

Figure 2: Confusion Matrix of Naïve Bayes Algorithm

Figure 2 presents the confusion matrix of the Naïve Bayes classifier used for fraud transaction detection. The matrix compares actual and predicted transaction classes, showing the numbers of correctly and incorrectly classified instances. Most fraudulent transactions were misclassified, indicating lower detection capability. This result highlights the limitations of the Naïve Bayes algorithm in handling highly imbalanced financial transaction datasets.

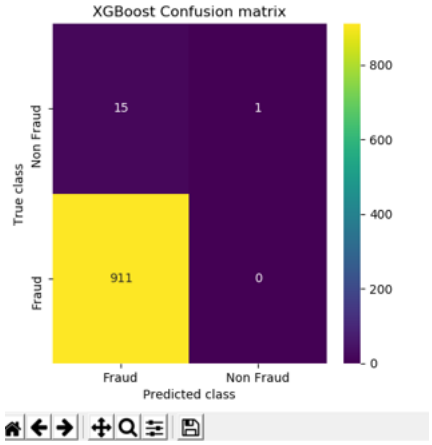

Figure 3: Confusion Matrix of XGBoost Algorithm

Figure 3 illustrates the confusion matrix of the proposed XGBoost classifier for fraud transaction detection. The matrix compares the actual transaction labels with the predicted outcomes, showing improved classification performance over the Naïve Bayes model. XGBoost correctly identifies a larger number of fraudulent transactions while reducing misclassifications, demonstrating its effectiveness in detecting fraud within highly imbalanced financial transaction datasets.

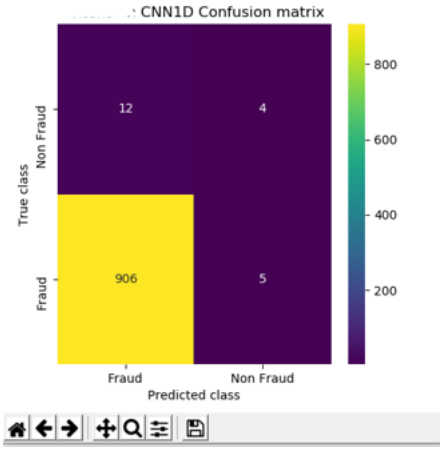

Figure 4: Confusion Matrix of CNN1D Algorithm

Figure 4 presents the confusion matrix of the CNN1D model for fraud transaction detection. The matrix compares actual transaction classes with predicted results, illustrating the model's classification performance on the testing dataset. CNN1D effectively captures complex transaction patterns and correctly identifies most fraudulent transactions while minimizing classification errors. This demonstrates the capability of deep learning for accurate and reliable fraud detection.

DISCUSSION

The experimental results demonstrate that the proposed hybrid fraud detection framework effectively improves the identification of fraudulent online financial transactions. The comparison of Naïve Bayes, XGBoost, and CNN1D shows that advanced learning techniques provide better predictive performance than traditional classification methods. Data preprocessing and feature selection using Principal Component Analysis (PCA) reduced redundant information and enhanced the quality of the input data, contributing

to improved model efficiency. XGBoost achieved higher accuracy by capturing complex relationships among transaction features, while CNN1D further enhanced detection by learning hidden patterns from the processed data. Evaluation using accuracy, precision, recall, F1-score, and confusion matrix confirmed that the proposed approach reduced misclassification and false predictions compared to the baseline model. The results indicate that integrating machine learning and deep learning techniques creates a reliable and scalable fraud detection system capable of handling large volumes of financial transactions. Therefore, the proposed framework offers an effective solution for strengthening the security of modern digital payment systems and supporting financial institutions in detecting fraudulent activities more accurately.

VI.CONCLUSION

The proposed hybrid fraud detection system provides an effective approach for identifying fraudulent online financial transactions by combining Principal Component Analysis (PCA), XGBoost, and CNN1D**.** The preprocessing stage improves data quality through normalization and feature selection, enabling the models to learn from the most relevant transaction attributes. Comparative analysis with the existing Naïve Bayes algorithm demonstrates that the proposed methods achieve better performance across evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix. XGBoost efficiently captures complex relationships within transaction data, while CNN1D further enhances fraud detection by learning hidden patterns and reducing misclassification. The experimental results confirm that the hybrid framework offers reliable and accurate classification of legitimate and fraudulent transactions, making it suitable for real-world financial applications. Overall, the proposed system strengthens the security of digital payment platforms by providing a scalable, intelligent, and efficient solution for fraud detection. Future enhancements may include real-time transaction monitoring, integration with streaming data platforms, and the use of advanced deep learning architectures to further improve detection accuracy and adaptability to emerging fraud patterns.

REFERENCES

1. R. Bhattacharyya and S. Jha, "Credit Card Fraud Detection using Machine Learning Algorithms", International Journal of Computer Applications, vol. 175, no. 12, pp. 23–29, 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, vol. 13, no. 1, pp. 34–42, 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]. Available: https://scikit-learn.org/stable/modules/decomposition.html

10.Kaggle, "Credit Card Fraud Detection Dataset", [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud