Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Performance Analysis of Machine Learning Techniques in Credit Card Fraud Detection

Suganya I[1], Naveen K.[2], Ragul P.[3], Sangeeth M.[4]

[1]Assistant Professor, [2-4]Student, Muthayammal Engineering College, Namakkal, Tamilnadu, India

E-mail: [1]inkarsalsuganya@gmail.com, [2]naveenkaasi@gmail.com, [3]ragulragul7358@gmail.com, [4]msangeeth109@gmail.com

Abstract

The rapid growth of e-commerce and online banking has resulted in a substantial rise in credit card fraud incidents. Consequently, machine learning and advanced deep learning techniques have emerged as critical solutions. This study integrates the findings of a few researchers, examining diverse methodologies, including Naïve Bayes, K-Nearest Neighbor (KNN), Logistic Regression, CNN, RNN, and ensemble learning. A comparative performance analysis, emphasizing the challenges posed by imbalanced datasets, demonstrates the superior performance of hybrid models, in enhancing the accuracy of detecting the fraudulent in credit card transactions.

Keywords: Credit Card Fraud Detection, Machine Learning, deep learning, Online Banking Security, Fraud Prevention.

1. Introduction

As the digital economy continues to grow, credit card fraud has emerged as a significant challenge, threatening both financial institutions and customers. Traditional fraud detection systems rely heavily on rule-based methods and statistical models, which, while effective to some extent, often fail to adapt to the dynamic and evolving nature of fraudulent behaviours. These systems are frequently unable to detect sophisticated fraud patterns, leading to false positives and negatives that affect their effectiveness.

Moreover, the increasing volume of transactional data in real-time scenarios necessitates the use of advanced computational approaches. Many existing systems struggle to

DOI: https://doi.org/10.36548/jscp.2024.4.005
Received: 16.12.2024, received in revised form: 07.01.2025, accepted: 28.01.2025, published: 07.02.2025

process large-scale data efficiently, thereby limiting their applicability in real-time fraud prevention. The problem is further compounded by the inherent imbalance in datasets, where legitimate transactions significantly outnumber fraudulent ones, making it challenging for models to accurately distinguish between the two categories.

To address these challenges, many machine learning and deep learning-based framework that enhances credit card fraud detection through intelligent data preprocessing, advanced classification algorithms, and real-time processing capabilities are being used. The proposed study is a comparative study of various machine learning and deep learning models that are used in addressing the challenges of traditional fraud detection methods. The study contributions are as follows

- To compare various research studies on credit card fraud detection using machine learning and deep learning techniques.
- To identify the challenges in existing methods.
- To suggest a hybridized model to improve the accuracy of identifying credit card fraud.

2. Literature Survey

The increasing prevalence of credit card fraud, driven by the growth of online transactions, has paved the way for extensive research into advanced fraud detection systems. Traditional methods, such as rule-based systems and statistical analysis, provide limited adaptability in detecting the evolving patterns of fraudulent behaviour. Moreover, these approaches often result in high false-positive rates, disrupting legitimate transactions and undermining user trust. These shortcomings highlight the need for robust, intelligent systems capable of identifying fraud with high accuracy while maintaining efficiency and scalability.

Recent advancements in machine learning have offered promising avenues for enhancing fraud detection. For example, researchers have explored supervised learning algorithms like Decision Trees and Support Vector Machines, which classify transactions as legitimate or fraudulent based on historical patterns, publicly available datasets, real-time datasets generated, and synthetic datasets. Table 1 below summarizes the methods and their performance in detecting credit card fraud.

Table 1. Comparative Study

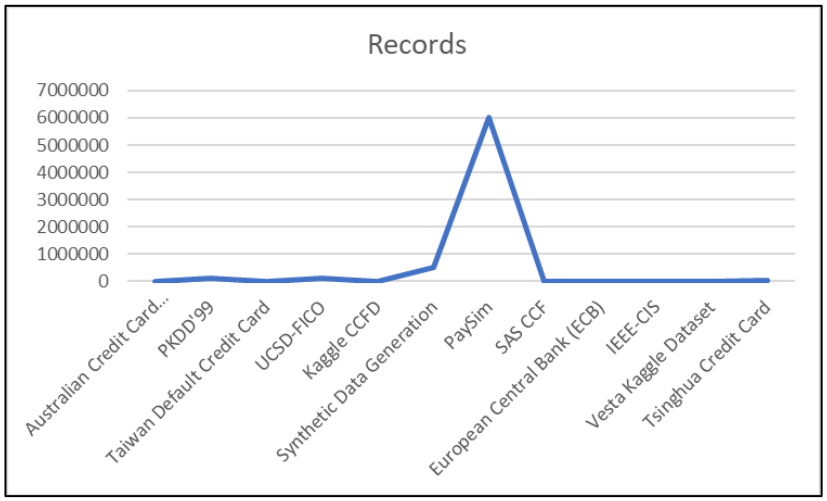
Methods Used	Dataset Used	Focus of the Study	Performance	Merits	Demerits
Federated Learning (FedFusion), Multilayer Perceptron (MLP) [1]	Three distinct datasets (1. Credit card transactions of European card holders, total 284,807, 2. Credit card fraud dataset from Kaggle, 3. Synthetic dataset from Kaggle	Secure and private training in fraud detection with heterogeneous data	Dataset1: 99.74%, Dataset2: 99.70%, and Dataset3: 96.61% detection rates	Addresses data heterogeneity, effective fraud detection	Faces implementation complexity, model convergence issues
Naïve Bayes, k-NN, Logistic Regression [2]	European cardholders (284,807 transactions)	Effect of sampling and variable selection on fraud detection	Naïve Bayes: 97.92%, k-NN: 97.69%, Logistic Regression: 54.86% accuracy	Naïve Bayes performed best among all tested models and Logistic Regression was the worst.	The biased dataset affects performance
Behavioral analysis, window-sliding strategy, classifier set training [3]	Real-time cardholder transactions	Fraud detection based on behavioral patterns	~80% accuracy, improved recall and accuracy	Adaptive to behavioral patterns, address concept drift	Shows varying performance across metrics
ML (Extreme Learning Method, Decision Tree, Random Forest, SVM, Logistic Regression,	European card benchmark dataset	Comparison of ML and DL techniques for fraud detection	Accuracy: 99.9%, F1-score: 85.71%, Precision: 93%, AUC: 98%	CNN outperformed ML models, reduced false negatives	Computationally intensive

XGBoost), Deep Learning (CNN) [4]					
ReMEMBeR (Ranking Metric Embedding-Based Multi-Contextual Behavior Profiling), Collaborative Filtering [5]	Real-world online banking dataset (3,462,533 transactions)	Context-aware fraud detection using recommender system techniques	Accuracy: 99.57%, Precision: 0.8642, Recall: 0.7620, F1-score: 0.8099	Effective for biased datasets, adaptable fraud profiling	Showed high false positive rate
VAEGAN (Variational Autoencoder GAN), Oversampling [6]	Kaggle credit card fraud dataset (European cardholders)	Improved fraud detection in imbalanced datasets using enhanced oversampling	F1-score: 0.884, Precision increased by 0.0203	Improved recall and specificity, better than SMOTE	GAN-based oversampling performance is instable.
Mahalanobis Distance SMOTE-ENN Hybrid Sampling, Random Forest [7]	Kaggle dataset, Taiwan credit card customer dataset	Fraud detection through hybrid sampling and RF	Recall: 91%, Precision: 93.66%, F1-score: 92%, AUC: 94%	Effective hybrid sampling, improved fraud detection	High Computational cost
Optimized Deep Event-Based Network (OptDevNet), ML classifiers [8]	Credit Card Fraud Detection (CCFD) Dataset	Deep learning model optimization for fraud detection	99.8% accuracy with fewer training iterations	High detection accuracy with fewer epochs	Limited dataset generalization
Deep Neural Networks [9]	Fraud detection system alerts (	Reducing false positives in	91.79% fraud detection with 35.16%	Cost reduction in the human review process	Some fraud cases required

	real-time dataset)	fraud detection	alert reduction		human validation
ResNeXt-Embedded GRU (RXT), Jaya Optimization, SMOTE, Autoencoders [10]	IEEE-CIS, Paysim, UCI Credit Card Dataset	Real-time fraud detection using AI	10%-18% improvement over existing methods	Computational efficiency, real-time fraud detection	Required extensive feature engineering
Multi-Layer Perceptron (MLP) [11]	Unspecified dataset	Optimization of activation functions for fraud detection	Sensitivity: 82%-83% depending on node count	Provides insight into MLP parameter tuning	Limited dataset details
Deep Neural Networks, One-Hot Encoding [12]	Real-world transactions (151,114 rows)	Identity fraud detection based on user behavior	Accuracy: 95%	Captures trading patterns for fraud detection	Limited dataset

While these methods improve detection accuracy, they often struggle with imbalanced datasets, where legitimate transactions vastly outnumber fraudulent ones. Hafez et al. [13], in his review, presents details of benchmark datasets introduced between 1992 and 2020, as well as recently available datasets introduced in 2023, which are commonly used in credit card fraud detection today. The author states that although these datasets enhance the detection accuracy of algorithms by providing details of complex fraud patterns, and real-world features including behavioral attributes and real-world fraud tactics, they still struggle with class imbalance. This necessitates efficient preprocessing and sampling techniques to mitigate the risk of overfitting and high processing demands.

The graph below illustrates the details of both benchmark and recent datasets, including the total number of records in each. Figure 1 depicts these datasets in more detail.



(a)

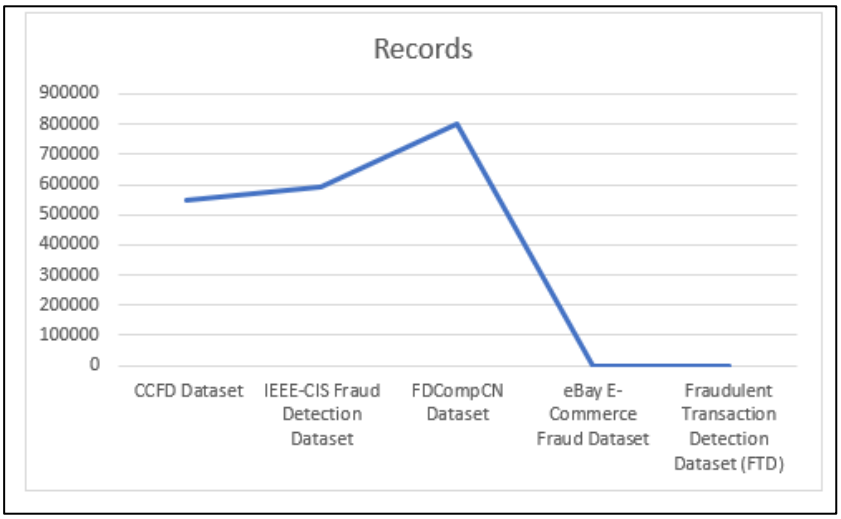


Figure 1. (a) and **(b)** Datasets for Credit Card Fraud Detection [13]

The author suggests that Hybrid sampling techniques, such as Synthetic Minority Oversampling (SMOTE), or ensemble methods should be been employed to address this issue by generating synthetic samples to balance the dataset. However, these methods require careful parameter tuning to prevent overfitting and ensure generalizability.

Real-time fraud detection is a critical requirement, especially in financial applications. Traditional batch-processing methods, while accurate, are often too slow to meet the demands of instantaneous fraud prevention. Recent studies propose integrating streaming analytics to ensure the timely detection of fraudulent activity. However, this approach requires high computational efficiency and resource optimization.

Another emerging concern is the need for privacy-aware systems, given the sensitive nature of financial data. Many existing fraud detection models rely on centralized processing, which increases the risk of data breaches. To address this, recent research has introduced privacy-preserving frameworks that process transactional data locally, minimizing exposure to unauthorized access. For example, Federated Learning enables decentralized model training across multiple devices without sharing raw data, enhancing privacy while maintaining model performance.

To comprehensively tackle these challenges, the proposed framework integrates advanced machine learning techniques to enhance credit card fraud detection. It employs hybrid sampling methods to address data imbalance and utilizes ensemble learning algorithms, such as Random Forest and Gradient Boosting, to improve classification accuracy. Real-time analytics enables the swift identification of suspicious transactions, while privacy-preserving mechanisms safeguard user data.

The suggested approach aims to balance accuracy, efficiency, and privacy within a robust fraud detection framework, laying the foundation for future advancements in securing financial transactions.

In summary, the suggested framework aims to address key challenges in credit card fraud detection accuracy, efficiency, scalability, adaptability, privacy, and transparency. Future work will focus on integrating state-of-the-art machine learning techniques with explainability and privacy-preserving mechanisms, this system sets a benchmark for intelligent, user-centric fraud detection solutions.

3. Proposed System Architecture

The proposed system enhances credit card fraud detection by integrating machine learning techniques and real-time data analysis to identify fraudulent transactions efficiently. It begins with a Transaction Data Collection Module, which collects real-time transaction data, including user details, transaction amounts, and merchant information. The Preprocessing Module applies data normalization and feature extraction techniques to prepare the transaction data for analysis, addressing data imbalances using hybrid sampling methods. The Hybrid Sampling Module uses SMOTE (Synthetic Minority Oversampling Technique) to address class imbalance, ensuring accurate detection of fraudulent transactions despite their lower frequency.

The classification Module utilizes multiple algorithms, such as Random Forest and Gradient Boosting, CNN, SVM, AdaBoost etc to increase classification accuracy while maintaining efficiency in real-time environments. Figure 2 depicts the general workflow of the credit card fraud detection system.

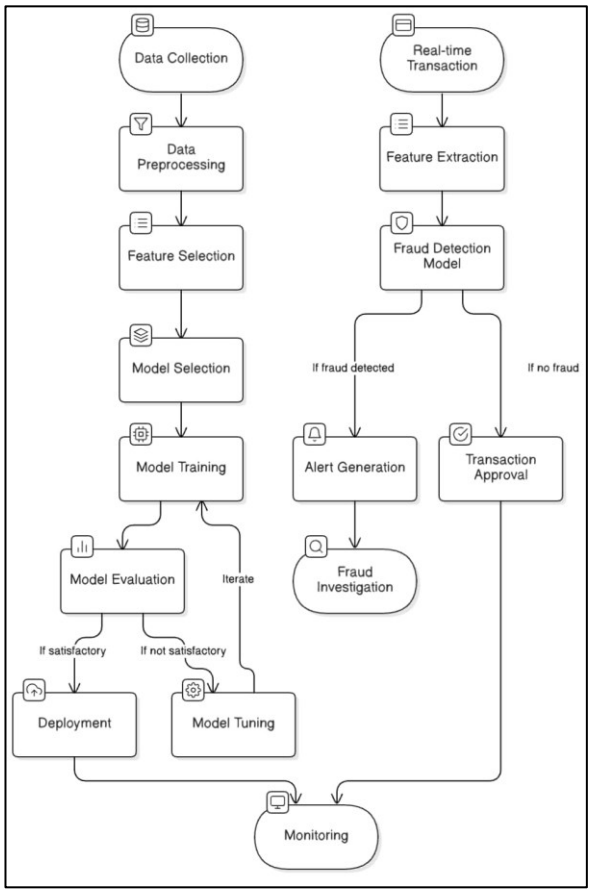


Figure 2. General Workflow of Credit Card Fraud Detection System.

4. Methodology

This study focuses on enhancing credit card fraud detection techniques by implementing and comparing multiple machine learning algorithms. The following methods were employed:

Several machine learning models were implemented to enhance credit card fraud detection. These include Logistic Regression (LR), which serves as a baseline for binary classification by distinguishing fraudulent and legitimate transactions using key features such as transaction amount and time; Random Forest (RF), an ensemble method that improves classification accuracy by constructing multiple decision trees, effectively handling imbalanced

datasets and capturing non-linear relationships; Gradient Boosting (XGBoost), which optimizes loss functions and reduces overfitting, making it ideal for handling sparse datasets and missing values; Convolutional Neural Networks (CNN), which can be trained on multimodal datasets, using both structured and unstructured features to improve detection accuracy; Adaptive Boosting (AdaBoost), which enhances fraud detection, particularly for minority fraud cases, by iteratively adjusting model weights to minimize classification errors; and finally, Support Vector Machines (SVM), which classify transactions by identifying optimal hyperplanes, with the radial basis function (RBF) kernel enabling the detection of complex fraud patterns in high-dimensional data. The Figure .3 below depicts the general stages in preparing the dataset

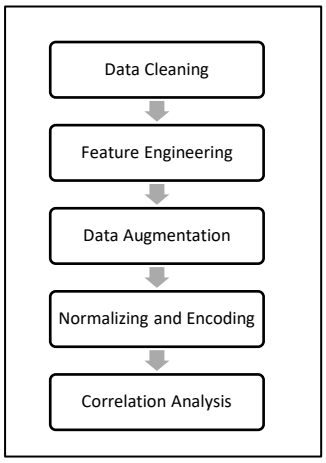


Figure 3. Dataset Preparation

The suggested system is expected to provide a significant improvement in credit card fraud detection through the integration of machine learning techniques and real-time data analysis. By using a robust data collection and preprocessing pipeline, the system is expected to efficiently handle real-time transaction data, ensuring it is properly normalized and feature-engineered for accurate analysis. The hybrid sampling approach, using SMOTE, addresses the challenge of class imbalance, which is crucial in detecting fraudulent transactions that are relatively rare. As a result, the system is expected to identify fraudulent activities with greater precision and accuracy. The use of multiple machine learning algorithms, including Random Forest, Gradient Boosting, CNN, AdaBoost, and SVM, is expected to achieve higher classification accuracy while ensuring it remains efficient in real-time environments. Each model brings unique advantages, such as handling missing data, capturing non-linear patterns, and adapting to imbalanced datasets, which collectively improve the overall detection performance. The incorporation of these models will lead to a more comprehensive and reliable

fraud detection system capable of minimizing false positives and improving the detection of genuine fraud cases.

5. Future work

This study aims to further evaluate the performance of the proposed model using a relevant dataset, along with appropriate preprocessing and sampling techniques, compare the effectiveness of each model, and ultimately deploy the optimal one in a user interface designed for real-time fraud detection

6. Conclusion

This study presents a brief overview of various machine learning and deep learning methods used for credit card fraud detection. It discusses the challenges affecting their performance and proposes a novel hybrid framework to enhance fraud detection by integrating advanced machine learning techniques with adaptive learning capabilities. The framework combines supervised and unsupervised algorithms, such as Random Forests, CNNs, and Autoencoders, to identify fraudulent transactions. Future work will focus on collecting an appropriate dataset, selecting suitable preprocessing, feature extraction, and sampling methods, and evaluating machine learning models to identify the optimal one. Finally, the selected model will be deployed in a user interface designed for real-time fraud detection.

References

[1] Nahid Ferdous Aurna, Md Delwar Hossain, Yuzo Taenaka,Youki Kadobayashi. "Adaptive Model Fusion for addressing Feature discrepancies in Federated Credit Card Fraud Detection." IEEE Access(2024).

[2] John O.Awoyemi , Adebayo O,Samuel. "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis." IEEE Access(2017).

[3] Changjun Jiang, Jiahui Song, Guanjun Liu, Member, IEEE, Lutao Zheng, and Wenjing Luan "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism." IEEE Access(2018).

[4] Fawaz Khaled Alarfaj , Iqra malik, Muhammad Ramzan , Muzamil Ahmed. "Credit Card Fraud Detection Using State-of the-Art Machine Learning and Deep Learning Algorithms." IEEE Access(2020).

[5] Jipeng Cui,Chungang Yan ,Cheng Wang."Ranking Metric Embedding Based Multicontextual Behavior Profiling for Online Banking Fraud Detection." IEEE Access(2021).

[6] Wei Kang, Yuanming Ding."Credit Card Fraud detection based on improved Variational Autoencoder Generative Adversial Network." IEEE Access(2023).

[7] Zhichao Xie, Xuan Huang."A Credit Card Fraud Detection Method Based Mahalanobis Distance Hybrid Sampling and Random Forest Algorithm." IEEE Access(2024).

[8] MuhammadAdil, Zhang Yinjun,Mona M. Jamjoom Zahid Ullah ."An Optimized Deep Event-Based Network Framework for Credit Card Fraud Detection." IEEE Access(2024).

[9] Rafael San Miguel Carrasco,Miguel Ángel Sicilia-Urbán. "Evaluation of Deep Neural Networks for Reduction of Credit Card Fraud Alerts." IEEE Access(2024).

[10] AbdulwahabAli Almazroi Nasir Ayub. "Online Payment Fraud Detection Model Using Machine Learning Techniques." IEEE Access(2024)

[11] Pillai, Thulasyammal Ramiah, Ibrahim Abaker Targio Hashem, Sarfraz Nawaz Brohi, Sukhminder Kaur, and Mohsen Marjani. "Credit card fraud detection using deep learning technique." In 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA) IEEE, 2018.1-6.

[12] Voican, Oona. "Credit Card Fraud Detection using Deep Learning Techniques." Informatica Economica 25, no. 1 (2021).

[13] Hafez, Ibrahim Y., Ahmed Y. Hafez, Ahmed Saleh, Abd El-Mageed, A. Amr, and Amr A. Abohany. "A systematic review of AI-enhanced techniques in credit card fraud detection." Journal of Big Data 12, no. 1 (2025): 1-35.