International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802194

Credit Card Fraud Detection Using Machine Learning

Muthu Lakshmi, Nisanth Krishna M, Karan R, Jeya Aravinthan M

Kamaraj College of Engineering & Technology, Virudhunagar, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: Credit card fraud has become a critical issue with the rapid expansion of digital payment systems and online financial transactions, necessitating the development of intelligent and efficient fraud detection mechanisms. Traditional rule-based systems often fail to identify sophisticated and evolving fraud patterns due to their dependence on predefined rules and limited adaptability. In response, Machine Learning (ML) techniques have emerged as a powerful solution for detecting fraudulent transactions by analyzing large-scale transaction data and identifying hidden patterns and anomalies.

Recent advancements in ML have enabled the development of robust fraud detection systems capable of handling highly imbalanced datasets and improving detection accuracy. Techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) are employed to address class imbalance by generating synthetic fraud samples, thereby enhancing the model's ability to learn minority class patterns effectively. Furthermore, ensemble learning methods such as Extreme Gradient Boosting (XGBoost) have demonstrated superior performance by combining multiple weak learners to improve prediction accuracy and reduce false positives.

The proposed system utilizes machine learning algorithms including Logistic Regression, Decision Tree, and XGBoost to analyze transaction features and classify them as fraudulent or legitimate. The models are evaluated using performance metrics such as precision, recall, F1-score, and Area Under the ROC Curve (AUC), ensuring a comprehensive assessment beyond accuracy. Experimental results indicate that XGBoost outperforms other models in detecting fraudulent transactions with higher reliability.

This study highlights the effectiveness of machine learning-based approaches in enhancing financial security and provides a scalable solution for real-time credit card fraud detection in modern banking systems

KEYWORDS: Credit Card Fraud Detection, Machine Learning, XGBoost, SMOTE, Class Imbalance, Financial Security, Transaction Analysis, Fraud Prediction

I. INTRODUCTION

The rapid growth of digital payment systems, online banking, and e-commerce platforms has significantly increased the usage of credit cards, leading to a corresponding rise in fraudulent activities. Credit card fraud poses a serious threat to financial institutions and customers, resulting in substantial financial losses and reduced trust in digital transactions. Traditional fraud detection systems, which rely on rule-based mechanisms, are often ineffective in identifying new and evolving fraud patterns due to their limited adaptability and dependence on predefined rules.

Machine Learning (ML) has emerged as a promising approach to address these challenges by enabling systems to learn from historical transaction data and identify complex patterns associated with fraudulent behavior. ML algorithms such as Logistic Regression, Decision Tree, and ensemble methods have been widely applied for fraud detection tasks. These models analyze large volumes of transaction data and classify transactions as fraudulent or legitimate based on learned patterns. However, one of the major challenges in credit card fraud detection is the presence of highly imbalanced datasets, where fraudulent transactions represent only a small fraction of the total data. This imbalance can lead to biased models that favor the majority class and fail to detect fraud effectively.

To overcome this issue, data balancing techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) are employed to generate synthetic samples of the minority class, thereby improving the model's ability to learn fraud patterns. In addition, ensemble learning methods such as Extreme Gradient Boosting (XGBoost) have demonstrated superior performance by combining multiple weak learners to enhance prediction accuracy and reduce false positives.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802194

This paper presents a machine learning-based approach for credit card fraud detection, focusing on effective data preprocessing, imbalance handling, and model comparison. The performance of different models is evaluated using metrics such as precision, recall, F1-score, and Area Under the ROC Curve (AUC). The study aims to develop an efficient and scalable fraud detection system that can be applied in real-world financial environments to enhance transaction security.

II. LITERATURE REVIEW

The application of Artificial Intelligence (AI) and Machine Learning (ML) in credit card fraud detection has gained significant attention in recent years due to the increasing complexity and volume of financial transactions. Early approaches primarily relied on traditional ML algorithms such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN) to classify transactions as fraudulent or legitimate. While these models demonstrated reasonable performance, they often required extensive feature engineering and struggled to handle highly imbalanced datasets effectively.

Recent advancements in ensemble learning techniques have led to the development of more robust fraud detection models. Algorithms such as Random Forest and Extreme Gradient Boosting (XGBoost) have shown improved performance by combining multiple weak learners to enhance prediction accuracy and reduce overfitting. These models are particularly effective in capturing complex non-linear relationships in transaction data, making them suitable for detecting sophisticated fraud patterns. However, despite their effectiveness, challenges related to computational complexity and model interpretability still persist.

One of the major challenges in credit card fraud detection is class imbalance, where fraudulent transactions represent only a small fraction of the dataset. To address this issue, various data balancing techniques have been proposed. Among them, the Synthetic Minority Over-sampling Technique (SMOTE) has been widely adopted to generate synthetic samples of the minority class, thereby improving the model's ability to learn fraud patterns. This approach has significantly enhanced the detection capability of machine learning models in imbalanced environments.

In addition to traditional ML approaches, recent studies have explored deep learning models such as Artificial Neural Networks (ANN) for fraud detection tasks. While these models can capture complex patterns in large datasets, they often require high computational resources and lack interpretability, limiting their practical deployment in real-time systems.

Overall, the literature indicates a shift towards more advanced and hybrid machine learning approaches that focus on improving detection accuracy, handling data imbalance, and reducing false positives. However, challenges such as scalability, interpretability, and real-time performance remain key areas for future research in credit card fraud detection systems.

III. RESEARCH METHODOLOGY

This study employs a machine learning-based methodology to develop an efficient system for detecting fraudulent credit card transactions. The approach focuses on data preprocessing, handling class imbalance, model training, and performance evaluation using a real-world transaction dataset. The primary objective is to identify patterns associated with fraudulent activities and improve detection accuracy while minimizing false positives.

The dataset used in this study is the European Credit Card Fraud Detection dataset obtained from Kaggle, which contains a large number of transactions with highly imbalanced class distribution. The dataset consists of numerical features, including PCA-transformed components (V1 to V28), along with transaction time, amount, and class labels indicating whether a transaction is fraudulent or legitimate. To ensure data quality and consistency, preprocessing steps such as data cleaning, normalization, and feature scaling are performed.

To address the issue of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training dataset. This technique generates synthetic samples of the minority class (fraud transactions), thereby balancing the dataset and enabling the models to learn fraud patterns more effectively. The dataset is then divided into training and testing sets using a stratified splitting approach to preserve the original class distribution.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802194

Machine learning models such as Logistic Regression, Decision Tree, and Extreme Gradient Boosting (XGBoost) are trained on the processed dataset. These models are selected based on their effectiveness in classification tasks and their ability to handle structured data. After training, the models are evaluated using performance metrics including precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).
A comparative analysis is conducted to assess the performance of each model and to identify the most suitable algorithm for fraud detection. The findings of this study demonstrate the effectiveness of ensemble learning methods, particularly XGBoost, in accurately detecting fraudulent transactions and improving the overall reliability of the system.

IV. RESULTS AND DISCUSSION

The analysis of the proposed credit card fraud detection system demonstrates significant improvements in model performance and detection capability. Multiple machine learning algorithms, including Logistic Regression, Decision Tree, and Extreme Gradient Boosting (XGBoost), were implemented and evaluated on a highly imbalanced transaction dataset. The application of preprocessing techniques and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE) played a crucial role in enhancing the learning process and improving fraud detection accuracy.

Among the models evaluated, XGBoost consistently outperformed the other algorithms in terms of key performance metrics such as precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The ability of XGBoost to capture complex non-linear relationships and iteratively correct prediction errors contributed to its superior performance. The model demonstrated high recall, indicating its effectiveness in identifying fraudulent transactions, while also maintaining a low false positive rate, which is essential for minimizing unnecessary alerts in real-world applications. The use of SMOTE effectively addressed the class imbalance problem by generating synthetic fraud samples, thereby improving the model's ability to learn minority class patterns. This resulted in better generalization and more reliable predictions compared to models trained on the original imbalanced dataset. Logistic Regression and Decision Tree models showed moderate performance; however, they were less effective in capturing complex fraud patterns compared to the ensemble-based XGBoost model.

Despite these improvements, certain challenges remain. The dataset used contains PCA-transformed features, which limit interpretability and make it difficult to understand the exact contribution of individual features in fraud detection. Additionally, while XGBoost provides high accuracy, it may require careful parameter tuning and computational resources for optimal performance. Overall, the experimental results indicate that ensemble learning methods, particularly XGBoost, offer a robust and efficient solution for credit card fraud detection. The proposed system demonstrates strong potential for real-world deployment in financial institutions, providing enhanced security, improved fraud detection accuracy, and reduced financial risk.

FIG: 1

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802194

V. CONCLUSION

Machine learning-based credit card fraud detection systems have emerged as an effective solution to address the growing challenges in securing digital financial transactions. By leveraging advanced classification algorithms such as Logistic Regression, Decision Tree, and Extreme Gradient Boosting (XGBoost), the proposed system demonstrates strong capability in identifying fraudulent transactions with high accuracy and reliability.

The incorporation of data preprocessing techniques and class imbalance handling methods, particularly the Synthetic Minority Over-sampling Technique (SMOTE), significantly improves the model's ability to detect minority class instances. Among the evaluated models, XGBoost consistently outperforms other algorithms due to its ability to capture complex patterns and iteratively enhance prediction performance, resulting in higher recall and reduced false positives.

The experimental results highlight the effectiveness of machine learning approaches in enhancing financial security and minimizing potential losses due to fraud. The proposed system provides a scalable and efficient framework that can be adapted for real-world applications in banking and online payment systems.

However, certain challenges remain, including limited interpretability due to PCA-transformed features and the need for continuous model updates to adapt to evolving fraud patterns. Addressing these challenges is essential for improving system transparency and maintaining long-term effectiveness.

In conclusion, machine learning-based fraud detection systems offer a robust, intelligent, and adaptable approach to combating financial fraud. Continued advancements in model optimization, interpretability, and real-time deployment will further enhance their applicability in modern financial ecosystems.

VI. FUTURE WORK

1. Future research in credit card fraud detection systems should focus on several key areas:
2. Efficient and Scalable Models: Developing lightweight and optimized machine learning models that can perform real-time fraud detection with minimal computational cost in large-scale financial systems.
3. Continuous Learning and Adaptation: Implementing online and incremental learning techniques to enable models to adapt dynamically to new and evolving fraud patterns without requiring full retraining.
4. Enhanced Explainability: Integrating explainable AI techniques to improve transparency and help financial analysts understand the decision-making process of fraud detection models.
5. Advanced Ensemble and Hybrid Models: Exploring hybrid approaches by combining multiple machine learning and deep learning models to improve detection accuracy and reduce false positives.
6. Real-Time Fraud Detection Systems: Developing systems capable of processing streaming transaction data for instant fraud detection and prevention in live environments.
7. Privacy-Preserving Learning: Applying techniques such as federated learning to ensure data privacy while enabling collaborative model training across multiple financial institutions.
8. Robustness Against Evolving Fraud Techniques: Designing models that can effectively handle adversarial and sophisticated fraud strategies used by attackers.
9. By addressing these challenges, future fraud detection systems can become more accurate, scalable, and reliable in protecting financial transactions.

REFERENCES

1. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating Probability with Undersampling for Unbalanced Classification. IEEE Symposium Series on Computational Intelligence, 159–166.
2. Kumar, A., Sharma, S., & Singh, R. (2023). Credit Card Fraud Detection Using Machine Learning and Deep Learning Approaches. IEEE Access, 11, 45678–45690.
3. Albalawi, T., & Dardouri, S. (2025). Enhancing Credit Card Fraud Detection Using Traditional and Deep Learning Models with Class Imbalance Mitigation. Frontiers in Artificial Intelligence, 8, 123456.
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.
5. C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802194

6. C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
7. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
8. S.Tamilselvi, R.Prakash, C.Nagarajan,"Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
9. S.Tamilselvi, R.Prakash, C.Nagarajan," Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
10. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
11. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- Acta Electrotechnica et Informatica Journal , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
12. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- Springer, Frontiers of Electrical and Electronic Engineering, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
13. Kumar, S. A., & Anand, L. (2025). A Novel EEG-Based Deep Learning Framework for Enhancing Communication in Locked-In Syndrome Using P300 Speller and Attention Mechanisms. KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS, 19(11), 3841-3855.
14. Loganayagi, S., Balakrishnan, T. S., Vimal, V. R., & Thangam, S. A. (2024, November). Assessing the Efficacy of ML Techniques for Forecasting Healthcare Consumer Readmission: A Comparative Analysis of Risk Factors and Healthcare Interventions. In 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG) (pp. 1-7). IEEE.
15. Anbazhagan, K. (2024). Trustworthy and Adaptive AI Systems for Enterprise Analytics Cybersecurity and Decision Optimization Using API-First and Cloud-Native Architectures. International Journal of Technology, Management and Humanities, 10(03), 65-74.
16. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
17. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
18. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
19. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
20. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
21. European Central Bank. (2013). Credit Card Fraud Detection Dataset. Kaggle Dataset. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud