Adaptive Machine Learning Approach for Banking Transaction Fraud Identification

Posam Srilakshmi[1], Dr. G. Purna Chandar Rao[2]
[1]MCA Student , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad
[2]Assistant Professor , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad

To Cite this Article
Posam Srilakshmi, Dr. G. Purna Chandar Rao, "Adaptive Machine Learning Approach for Banking Transaction Fraud Identification", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 07, July 2026, pp: 964-970, DOI: http://doi.org/10.64771/jsetms.2026.v03.i07.pp964-970
Submitted: 19-06-2026 Accepted: 24-07-2026 Published: 30-07-2026

Abstract- Financial fraud has become one of the major challenges faced by banking and digital payment systems due to the increasing volume of online transactions. Detecting fraudulent activities at an early stage is essential for minimizing financial losses and improving the security of banking services. This paper presents a machine learning-based fraud detection system that identifies suspicious banking transactions by analyzing historical transaction records. Before model training, the collected dataset undergoes preprocessing to eliminate missing values, remove non-numeric attributes, and transform the data into a suitable format for classification. Several supervised learning algorithms, including Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), Decision Tree, AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), and Random Forest, are implemented and evaluated using performance measures such as accuracy, precision, recall, and F1-score. Experimental analysis shows that the Random Forest classifier outperforms the remaining models by providing the highest classification accuracy and more reliable fraud prediction. The developed application offers an interactive interface for dataset preprocessing, model training, algorithm comparison, and fraud prediction, enabling users to analyze transaction data efficiently. The comparative evaluation demonstrates that ensemble learning techniques can significantly improve the detection of fraudulent banking transactions while reducing false classifications. The proposed framework provides an effective, scalable, and practical solution for enhancing fraud detection capabilities in modern banking systems and supporting secure digital financial transactions.

Keywords— Fraud Detection, Banking Security, Machine Learning (ML), Random Forest, Support Vector Machine (SVM), Logistic Regression, Decision Tree, AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), Transaction Classification, Fraud Prediction, Banking Transactions, Blockchain Transactions, Supervised Learning, Data Preprocessing, Financial Fraud Detection, Classification Algorithms, Intelligent Banking System, Cybersecurity.

I.INTRODUCTION

The rapid expansion of digital banking, online payment platforms, and blockchain-based financial services has significantly increased the number of electronic transactions processed every day. While these technologies provide greater convenience, speed, and accessibility, they have also created new opportunities for fraudulent activities that threaten the security and reliability of financial systems **[3]**. Fraudulent transactions can result in substantial financial losses for banks, businesses, and customers, making fraud detection one of the most important challenges in modern banking environments. Consequently, financial institutions are increasingly adopting intelligent data-driven techniques to identify suspicious transactions before they cause economic damage **[7]**.
Traditional fraud detection approaches are primarily based on manually defined rules and expert knowledge. Although these methods are effective in identifying known fraudulent patterns, they often struggle to detect newly emerging fraud strategies and require continuous updates as transaction behavior changes **[9]**. In addition, the growing volume of financial data makes manual analysis both time-consuming and inefficient. These limitations have encouraged researchers to investigate machine learning techniques that automatically learn transaction patterns from historical data and accurately distinguish between legitimate and fraudulent activities **[10]**.
Recent developments in machine learning have enabled the design of intelligent fraud detection systems capable of processing large-scale transaction datasets with improved accuracy and efficiency. Supervised learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), Naïve Bayes,

AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Networks (DNN), and Random Forest, have been widely applied for binary classification problems such as fraud detection **[13]**. These algorithms analyze historical transaction records, identify hidden relationships among transaction attributes, and classify each transaction as either genuine or fraudulent. Proper data preprocessing, including handling missing values, removing irrelevant information, and converting non-numeric attributes into numerical representations, further improves the performance and reliability of these classification models **[15]**.

The proposed work presents a machine learning-based fraud detection system for banking transactions that analyzes transaction data and predicts fraudulent activities through a user-friendly application. The system performs dataset preprocessing, generates training and testing datasets, and evaluates multiple machine learning algorithms using standard performance metrics such as accuracy, precision, recall, and F1-score **[4]**. A comparative analysis is carried out to determine the most effective classification model for fraud prediction. The application also provides facilities for visualizing algorithm performance and predicting the fraud status of new transactions using the trained model **[8]**.

Experimental evaluation demonstrates that the Random Forest classifier produces better prediction performance than the other evaluated algorithms for the banking transaction dataset considered in this study **[12]**. The comparative results indicate that ensemble learning methods can effectively capture complex transaction patterns and improve fraud detection accuracy while reducing misclassification. The proposed system offers a practical and scalable solution for strengthening banking security, supporting financial institutions in minimizing fraudulent transactions, and improving the overall reliability of digital banking services **[18]**.

II.RELATED WORK

J. Nanduri, Y.-W. Liu, K. Yang, and Y. Jia proposed a machine learning framework for detecting fraudulent activities in e-commerce transactions by identifying groups of related fraudulent accounts, referred to as fraud islands, and combining them with a multi-layer learning model [1]. The study demonstrated that analyzing relationships among transaction records improves the ability to recognize fraudulent behavior that may not be detected through traditional classification techniques alone. The proposed framework enhanced detection accuracy by integrating multiple levels of analysis, making it suitable for large-scale online financial systems. Their work highlighted the importance of combining structural transaction information with machine learning models to strengthen fraud detection and improve the overall security of digital payment platforms [1].

K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi introduced a credit card fraud detection approach based on the AdaBoost algorithm combined with a majority voting strategy to improve classification performance [5]. The proposed method utilized the strengths of multiple classifiers to reduce incorrect predictions and increase the reliability of fraud identification. Experimental results showed that ensemble learning techniques were more effective than individual classifiers in distinguishing fraudulent transactions from legitimate ones. The authors concluded that combining boosting methods with voting mechanisms provides a practical solution for developing accurate and dependable fraud detection systems for financial institutions [5].

C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten investigated the impact of feature engineering on credit card fraud detection using machine learning models [10]. Their research emphasized that carefully selecting and transforming transaction attributes significantly improves the predictive performance of classification algorithms. The study demonstrated that meaningful feature extraction enables machine learning models to capture hidden transaction patterns more effectively, resulting in higher fraud detection accuracy. The authors highlighted feature engineering as an essential step in building robust fraud detection systems capable of handling complex financial datasets [10].

F. Itoo, M. Meenakshi, and S. Singh presented a comparative study of Logistic Regression, Naïve Bayes, and K-Nearest Neighbors (KNN) algorithms for detecting fraudulent credit card transactions [13]. The research evaluated the strengths and limitations of each classifier using standard performance measures to determine their suitability for fraud detection tasks. The experimental findings indicated that classification accuracy varies depending on the characteristics of the transaction dataset and the learning algorithm employed. The study provided valuable insights into selecting appropriate supervised learning techniques for financial fraud prediction and highlighted the importance of comparative evaluation before deploying machine learning models in real-world applications [13].

H. Wang, P. Zhu, X. Zou, and S. Qin developed an ensemble learning framework that improves credit card fraud detection by partitioning the training dataset and applying clustering techniques before model training [12]. The proposed approach was designed to address the challenges associated with highly imbalanced financial datasets and to improve the learning capability of classification models. By combining multiple learning strategies, the framework achieved better prediction performance and reduced false classifications compared to conventional methods. The authors demonstrated that ensemble-based approaches provide an effective and scalable solution for identifying fraudulent financial transactions in modern banking environments [12].

III.DATASET DETAILS

The Fraud Detection in Banking Data using Machine Learning Techniques system is developed using a structured transaction dataset containing records of banking and blockchain-based financial transactions. Each record represents the details of an individual transaction along with its corresponding class label indicating whether the transaction is legitimate or fraudulent. The dataset includes various transaction-related attributes that describe user activity and transaction characteristics, enabling machine learning algorithms to learn patterns associated with fraudulent behavior. Before model development, the dataset undergoes several preprocessing steps, including handling missing values, removing non-numeric attributes, and converting categorical information into numerical form suitable for machine learning models. The target attribute, **FLAG**, is used as the class label, where a value of **0** represents a normal transaction and **1** represents a fraudulent transaction. After preprocessing, the dataset is divided into training and testing subsets for model development and performance evaluation. The processed dataset serves as the foundation for training multiple supervised learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), and Random Forest, allowing a comparative analysis of their fraud detection capabilities. During system execution, the trained models analyze transaction records to determine whether a new transaction is genuine or fraudulent. Incoming transaction data is first preprocessed using the same transformation techniques applied during model training to ensure consistency in prediction. The extracted numerical features are then provided to the trained classifiers, which estimate the probability of fraud based on patterns learned from historical transaction data. The application also provides facilities for uploading datasets, generating training and testing data, comparing the performance of multiple machine learning algorithms, and displaying evaluation metrics such as accuracy, precision, recall, and F1-score. Comparative experimental results demonstrate that the **Random Forest** classifier achieves the highest prediction performance among the evaluated models, making it the most suitable algorithm for the proposed fraud detection framework. The structured organization of transaction records and systematic preprocessing process enable the system to deliver accurate fraud prediction, support efficient transaction monitoring, and strengthen the security of modern banking and digital payment environments.

IV.PROPOSED METHODOLOGY

The proposed **Fraud Detection in Banking Data using Machine Learning Techniques** system is developed to accurately identify fraudulent banking transactions using supervised machine learning. The methodology consists of four main stages: data preprocessing, model training, model evaluation, and fraud prediction. Initially, the transaction dataset is uploaded into the system, where missing values are handled, non-numeric attributes are removed, and the remaining data is transformed into a numerical format suitable for machine learning algorithms. The target attribute, **FLAG**, is used to classify each transaction as either legitimate (0) or fraudulent (1). The processed dataset is then divided into training and testing sets for model development. Several machine learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), and Random Forest, are trained and evaluated using accuracy, precision, recall, and F1-score. Comparative analysis shows that the **Random Forest** classifier achieves the highest prediction accuracy among all evaluated models. Random Forest constructs multiple decision trees using different subsets of the training data and predicts the final class through a majority voting mechanism, thereby improving robustness and reducing overfitting. During system execution, every new transaction undergoes the same preprocessing steps before being analyzed by the trained Random Forest model. The predicted result is displayed as either a normal or fraudulent transaction, providing a reliable, scalable, and efficient solution for securing modern banking systems against financial fraud.

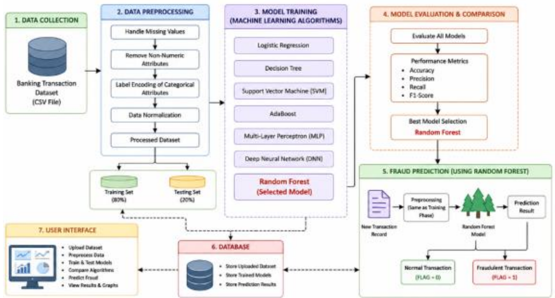

Figure 1: System Architecture of the Proposed Fraud Detection in Banking Data Using Random Forest

Figure 1 illustrates the architecture of the proposed fraud detection system, which combines data preprocessing, machine learning, and prediction to identify fraudulent banking transactions. Initially, the banking transaction dataset is uploaded and preprocessed by handling missing values, removing non-numeric attributes, encoding categorical data, and normalizing the dataset. The processed data is then divided into training and testing sets. Several supervised learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), and Random Forest, are trained and evaluated using accuracy, precision, recall, and F1-score. Based on comparative analysis, Random Forest is selected as the final prediction model because it achieves the best performance. During prediction, new transaction records undergo the same preprocessing steps before being analyzed by the trained Random Forest classifier. The model classifies each transaction as either normal (FLAG = 0) or fraudulent (FLAG = 1). The system also stores datasets, trained models, and prediction results while providing an interface for dataset upload, model comparison, fraud prediction, and result visualization.

V.RESULT AND DISCUSSION

The proposed Fraud Detection in Banking Data using Machine Learning Techniques system was successfully implemented and evaluated using a machine learning-based application for identifying fraudulent banking transactions. The developed system effectively performed all major operations, including dataset upload, data preprocessing, training and testing of machine learning models, fraud prediction, algorithm comparison, and graphical visualization of performance metrics. During experimentation, the banking transaction dataset was cleaned by handling missing values, removing non-numeric attributes, and converting the remaining features into a suitable numerical format for classification. Multiple supervised learning algorithms, namely Logistic Regression, Naïve Bayes, Decision Tree, Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), and Random Forest, were trained and evaluated using accuracy, precision, recall, and F1-score. Comparative analysis revealed that the Random Forest classifier achieved the highest prediction accuracy among all evaluated models and was therefore selected as the final fraud detection model. The application successfully classified incoming transactions as either normal (FLAG = 0) or fraudulent (FLAG = 1) and displayed the prediction results through a simple user interface. The system also maintained the uploaded dataset, trained models, and prediction outcomes for future analysis and monitoring.

Experimental results demonstrated that the proposed framework accurately identified fraudulent transactions while maintaining consistent prediction performance across different transaction records. The Random Forest classifier effectively captured complex transaction patterns by combining the decisions of multiple decision trees through a majority voting mechanism, resulting in improved classification accuracy and reduced false predictions compared with the other evaluated algorithms. The performance comparison graphs further confirmed the superiority of the Random Forest model in terms of accuracy, precision, recall, and F1-score. The developed application provides an efficient environment for preprocessing transaction data, training machine learning models, comparing algorithm performance, and predicting fraud in real time. Overall, the proposed system offers a reliable, scalable, and practical solution for strengthening banking security, reducing financial fraud, and supporting secure digital banking operations through intelligent machine learning-based fraud detection.

Random Forest Accuracy : 94.1
Random Forest Precision : 93.61522535435579
Random Forest Recall : 89.13985502275274
Random Forest FScore : 91.12460305163225

Figure 2: Performance of the Random Forest Classifier

Figure 2 presents the performance metrics of the Random Forest classifier used in the proposed fraud detection system. The model achieved an accuracy of 94.10%, indicating that it correctly classified the majority of banking transactions. A precision of 93.62% demonstrates the classifier's effectiveness in identifying fraudulent transactions while minimizing false positive predictions. The model also obtained a recall of 89.14%, showing its capability to detect a high percentage of actual fraudulent transactions. Furthermore, the F1-score of 91.12% reflects a balanced performance between precision and recall, confirming the reliability of the classifier. These experimental results indicate that the Random Forest algorithm outperforms the other evaluated machine learning models and provides consistent fraud detection performance. Its ensemble learning mechanism, which combines the predictions of multiple decision trees through majority voting, improves classification accuracy and reduces overfitting. Therefore, the Random Forest classifier was selected as the final prediction model for the proposed banking fraud detection system due to its superior accuracy, robustness, and effectiveness in identifying fraudulent financial transactions.

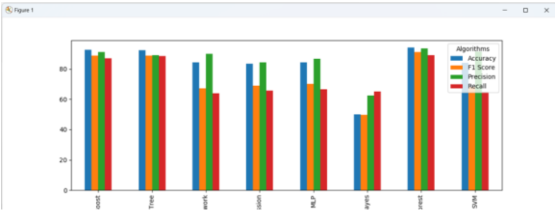

Figure 3: Comparative Performance Analysis of Machine Learning Algorithms

Figure 3 illustrates the comparative performance of the machine learning algorithms evaluated for fraud detection in banking transactions. The graph compares Accuracy, Precision, F1-score, and Recall for Logistic Regression, Multi-Layer Perceptron (MLP), Naïve Bayes, AdaBoost, Decision Tree, Support Vector Machine (SVM), Random Forest, and Deep Neural Network (DNN). The experimental results show that the Random Forest classifier achieved the highest overall performance, with an accuracy of 94.10%, precision of 93.62%, recall of 89.14%, and an F1-score of 91.12%. The Deep Neural Network also produced competitive results but performed slightly below Random Forest. AdaBoost and Decision Tree demonstrated satisfactory classification performance, whereas Logistic Regression, MLP, Naïve Bayes, and SVM achieved comparatively lower evaluation scores. The graph clearly indicates that the ensemble learning capability of Random Forest enables it to classify banking transactions more accurately while reducing misclassification. Based on these findings, the Random Forest algorithm was selected as the final prediction model for the proposed fraud detection system because it provides the most reliable and consistent performance across all evaluation metrics.

DISCUSSION

The proposed Fraud Detection in Banking Data using Machine Learning Techniques system provides an effective framework for identifying fraudulent financial transactions by integrating data preprocessing, supervised machine learning, and real-time prediction within a single application. The system supports the complete fraud detection workflow, including dataset upload, preprocessing, model training, algorithm comparison, fraud prediction, and performance visualization. A structured transaction dataset is used to train multiple machine learning models, while the application maintains processed datasets, trained models, and prediction results in a centralized repository for efficient management. Among the evaluated classifiers, Random Forest demonstrated the best overall performance, outperforming Logistic Regression, Decision Tree, Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), and Naïve Bayes. Its ensemble learning strategy, which combines the predictions of multiple decision trees through majority voting, improves classification accuracy and provides greater stability when analyzing complex banking transaction data. The user-friendly interface further enables straightforward dataset management, model evaluation, and fraud prediction without requiring extensive technical expertise.

The experimental implementation confirms that the proposed system successfully detects fraudulent transactions while maintaining consistent prediction performance across different transaction records. The preprocessing module effectively improves data quality by handling missing values, eliminating irrelevant attributes, and transforming the dataset into a format suitable for machine learning. This preprocessing stage contributes significantly to the overall effectiveness of the prediction models. Comparative analysis shows that the Random Forest classifier achieves higher accuracy, precision, recall, and F1-score than the other evaluated algorithms, making it the most suitable model for practical fraud detection. The developed framework provides a reliable, scalable, and cost-effective solution for strengthening banking security by identifying suspicious transactions quickly and accurately. Its ability to support efficient fraud monitoring, minimize false predictions, and adapt to large transaction datasets makes it well suited for deployment in modern banking and digital payment environments.

VI.CONCLUSION

The Fraud Detection in Banking Data using Machine Learning Techniques system provides an effective and dependable solution for identifying fraudulent banking transactions using supervised machine learning. The developed application integrates dataset preprocessing, model training, performance evaluation, and real-time fraud prediction into a single framework, enabling efficient analysis of financial transaction data. Several classification algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Deep Neural Network (DNN), Naïve Bayes, and Random Forest, were implemented and compared using standard evaluation metrics. Among these models, the Random Forest classifier achieved the highest prediction accuracy and demonstrated superior capability in distinguishing legitimate transactions from fraudulent ones. Its ensemble learning mechanism improved classification reliability by combining the predictions of multiple decision trees, thereby reducing false classifications and enhancing overall detection performance. The developed system also provides an intuitive interface for dataset upload, preprocessing, algorithm comparison, fraud prediction, and visualization of evaluation results, making it suitable for practical deployment in banking environments.

The experimental evaluation confirms that the proposed framework can accurately detect suspicious transactions while maintaining consistent performance across different transaction records. The preprocessing stage improves data quality by eliminating inconsistencies and preparing the dataset for effective model training, which contributes significantly to prediction accuracy. The scalable design of the system enables it to process large volumes of banking transaction data efficiently, supporting financial institutions in reducing fraud-related losses and improving transaction security. Future enhancements may include the integration of real-time transaction streaming, advanced deep learning models, explainable artificial intelligence techniques for transparent decision-making, cloud-based deployment for large-scale applications, adaptive learning from newly observed fraud patterns, and integration with banking information systems to further strengthen fraud detection capabilities and improve the security of digital financial services.

REFERENCES

[1] J. Nanduri, Y.-W. Liu, K. Yang, and Y. Jia, "E-commerce fraud detection through fraud islands and multi-layer machine learning model," in *Proc. Future Information and Communication Conference*, Advances in Information and Communication. Springer, 2020, pp. 556–570.

[2] H. Feng, "Ensemble learning in credit card fraud detection using boosting methods," in *Proc. 2nd International Conference on Computing and Data Science (CDS)*, Jan. 2021, pp. 7–11.

[3] M. S. Delgosha, N. Hajiheydari, and S. M. Fahimi, "Elucidation of big data analytics in banking: A four-stage Delphi study," *Journal of Enterprise Information Management*, vol. 34, no. 6, pp. 1577–1596, Nov. 2021.

[4] M. Puh and L. Brkić, "Detecting credit card fraud using selected machine learning algorithms," in *Proc. 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2019, pp. 1250–1255.

[5] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018.

[6] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, "Credit card fraud detection using a new hybrid machine learning architecture," *Mathematics*, vol. 10, no. 9, Art. no. 1480, Apr. 2022.

[7] K. Gupta, K. Singh, G. V. Singh, M. Hassan, G. Himani, and U. Sharma, "Machine learning based credit card fraud detection—A review," in *Proc. International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 2022, pp. 362–368.

[8] R. Almutairi, A. Godavarthi, A. R. Kotha, and E. Ceesay, "Analyzing credit card fraud detection based on machine learning models," in *Proc. IEEE International IoT, Electronics and Mechatronics Conference (IEMTRONICS)*, Jun. 2022, pp. 1–8.

[9] N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using artificial immune systems," *Applied Soft Computing*, vol. 24, pp. 40–49, Nov. 2014.

[10] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, Jun. 2016.

[11] U. Porwal and S. Mukund, "Credit card fraud detection in e-commerce: An outlier detection approach," *arXiv preprint arXiv:1811.02196*, 2018.

[12] H. Wang, P. Zhu, X. Zou, and S. Qin, "An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering," in *Proc. IEEE SmartWorld Congress*, Oct. 2018, pp. 94–98.

[13] F. Itoo, M. Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *International Journal of Information Technology*, vol. 13, no. 4, pp. 1503–1511, 2021.

[14] T. A. Olowookere and O. S. Adewale, "A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach," *Scientific African*, vol. 8, Art. no. e00464, Jul. 2020.

[15] A. A. Taha and S. J. Malebary, "An intelligent approach to credit card fraud detection using an optimized Light Gradient Boosting Machine," *IEEE Access*, vol. 8, pp. 25579–25587, 2020.

[16] X. Kewei, B. Peng, Y. Jiang, and T. Lu, "A hybrid deep learning model for online fraud detection," in *Proc. IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Jan. 2021, pp. 431–434.

[17] J. Zou, J. Zhang, and P. Jiang, "Credit card fraud detection using autoencoder neural network," *arXiv preprint arXiv:1908.11553*, 2019.

[18] M. Rakhshaninejad, M. Fathian, B. Amiri, and N. Yazdanjue, "An ensemble-based credit card fraud detection algorithm using an efficient voting strategy," *The Computer Journal*, vol. 65, no. 8, pp. 1998–2015, Aug. 2022.