International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

A New Robust Solution for Fraud Detection using Machine Learning

Vaibhav Sharma, Rohit Singh, Arijit Pasari, Sayandeep Saha, Shilpi Bose* and Chandra Das

[1]CSE Department
Netaji Subhash Engg. College
Kolkata-152, West Bengal, India.

***Abstract* - Credit card fraud detection remains a significant challenge due to the rapid growth of digital transactions and the highly imbalanced nature of financial datasets. This paper proposes a Bagging-Based Deep Learning Ensemble Framework to improve the accuracy and robustness of fraud detection. The framework generates multiple bootstrap samples from the training data and trains diverse deep learning models, including Autoencoder (AE), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Generative Adversarial Network (GAN), and Graph Neural Network (GNN). The best-performing model from each bootstrap is selected using the F1-score, and the selected models are combined through majority voting to produce the final prediction. The proposed framework is evaluated on both balanced and highly imbalanced credit card fraud datasets using 70:30 train-test split, 5-fold cross-validation, and 10-fold cross-validation. Experimental results demonstrate consistent performance across different validation strategies, highlighting the framework's ability to reduce prediction variance and improve detection reliability. The proposed approach offers a robust and generalized solution for real-world credit card fraud detection.**

Keywords - Credit Card Fraud Detection, Deep Learning, Ensemble Learning, Bootstrap Aggregating, Majority Voting, Autoencoder, CNN, LSTM, GAN.

I. INTRODUCTION

The rapid expansion of digital payment systems and online banking has transformed the global financial ecosystem by enabling fast, convenient, and secure financial transactions. However, this digital transformation has also increased the frequency and complexity of credit card fraud, resulting in substantial financial losses for banking institutions and consumers worldwide. According to recent reports, global payment card fraud losses continue to increase annually, making the development of intelligent fraud detection systems an important research challenge. Conventional rule-based detection techniques are no longer sufficient to identify evolving fraud patterns, leading researchers to explore machine learning and deep learning approaches for automated fraud detection [1, 2].

In recent years, numerous machine learning and deep learning techniques have been proposed for credit card fraud detection, including Support Vector Machines (SVM), Decision Trees, Random Forests, XGBoost, Autoencoders, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, Graph Neural Networks (GNN), and Generative Adversarial Networks (GAN) [1–4]. While these approaches have achieved promising results, most rely on a single learning model, making their performance sensitive to data distribution and class imbalance. Since fraudulent transactions generally constitute only a very small fraction of the available data, conventional models often exhibit high prediction variance, reduced stability, and poor generalization when evaluated under different validation strategies [2–4]. Consequently, improving robustness while maintaining reliable fraud detection remains a significant challenge.

To address these limitations, this paper proposes a Bagging-Based Deep Learning Ensemble Framework that combines the advantages of ensemble learning and multiple deep learning architectures for credit card fraud detection. The proposed framework generates multiple bootstrap samples from the training dataset and independently trains five deep learning models, namely Autoencoder (AE), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Generative Adversarial Network (GAN), and Graph Neural Network (GNN). For each bootstrap sample, the model with the highest F1-score is selected, and the final prediction is obtained through a majority voting mechanism using the selected models. This strategy aims to reduce prediction variance, improve model stability, and enhance the framework's ability to generalize across different data distributions.

The effectiveness of the proposed framework is evaluated using both balanced and highly imbalanced credit card fraud datasets under three validation strategies: 70:30 train-test split, 5-fold cross-validation, and 10-fold cross-validation. Experimental results demonstrate that the proposed ensemble framework provides consistent and reliable fraud detection performance, highlighting its suitability for real-world credit card fraud detection applications.

II. RELATED WORK

Credit card fraud detection has been extensively investigated using both machine learning and deep learning techniques due to the increasing volume of digital financial transactions and the evolving nature of fraudulent activities [1, 2]. Early studies primarily employed conventional machine learning algorithms such as Decision Trees, Support Vector Machines (SVM), Random Forests, Naïve Bayes, Logistic Regression, and XGBoost to classify legitimate and fraudulent transactions [1, 2, 4]. Although these methods achieved satisfactory classification performance, their effectiveness was often affected by severe class imbalance, resulting in reduced fraud detection capability and limited generalization across different datasets [2, 3].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

To improve detection performance, researchers have increasingly adopted deep learning approaches capable of learning complex and nonlinear transaction patterns. Autoencoders have been utilized for fraud detection and feature representation [3], while deep neural-network-based approaches have also demonstrated their applicability to credit card fraud detection [5]. Long Short-Term Memory (LSTM) networks have been investigated in fraud detection frameworks together with techniques for handling class imbalance [6]. More recently, hybrid machine learning and deep learning ensemble approaches have also been investigated to improve fraud detection performance [7]. Despite these advances, developing models that remain robust under different data distributions and severe class imbalance continues to be an important research problem.

Ensemble learning has emerged as an effective strategy for improving prediction accuracy and model stability by combining multiple classifiers [8–11]. Existing ensemble approaches include Stacking [8], Bagging [9], and Boosting [10], with each employing a different strategy for combining multiple learners. Ensemble methods can reduce dependence on an individual classifier and improve predictive robustness [11]. However, the existing fraud detection literature still presents challenges related to class imbalance, model stability, computational complexity, and comprehensive validation [2, 4, 7].

Despite these advances, achieving a fraud detection framework that simultaneously addresses class imbalance, prediction variance, and model generalization remains an important research challenge. Developing a robust ensemble approach capable of integrating complementary deep learning models and producing stable performance under diverse validation strategies is therefore of significant research interest.

Table 1. Summary of Existing Credit Card Fraud Detection Approaches

Reference	Methodology	Major Limitation
Ding et al. **[3]**	Autoencoder + LightGBM	Evaluated primarily using holdout validation
Farabi et al. **[4]**	Multiple Machine Learning Algorithms	Performance and stability depend on individual models
Btoush et al. **[7]**	CNN–BiLSTM Ensemble	High computational complexity
Wolpert **[8]**	Stacked Generalization	Additional computational complexity due to meta-learning
Breiman **[9]**	Bagging	Does not directly address severe class imbalance
Benchmark Study **[12]**	XGBoost	Lower recall for fraudulent transactions
Proposed Work	**Bagging-Based Deep Learning Ensemble (AE, CNN, LSTM, GAN, GNN)**	**Improved robustness through heterogeneous ensemble and multiple validation strategies**

III. PROPOSED METHODOLOGY

A. *Overview of the Proposed Framework*

This work proposes a Bagging-Based Deep Learning Ensemble Framework for credit card fraud detection to improve prediction accuracy, robustness, and generalization across different transaction distributions. Unlike conventional fraud detection methods that rely on a single classifier, the proposed framework integrates multiple deep learning models within a bagging ensemble to exploit their complementary learning capabilities.

The framework begins by dividing the dataset into training and independent holdout testing sets. Multiple bootstrap samples are then generated from the training data using sampling with replacement. Each bootstrap sample is independently used to train five deep learning models, namely Autoencoder (AE), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Generative Adversarial Network (GAN), and Graph Neural Network (GNN). The performance of these models is evaluated using multiple validation strategies, and the model achieving the highest F1-score within each bootstrap sample is selected as the winner. Finally, the predictions generated by all selected winner models on the independent holdout dataset are combined through a majority voting mechanism to obtain the final fraud prediction.

The proposed framework aims to reduce prediction variance, improve model stability, and enhance fraud detection performance by combining the strengths of multiple deep learning architectures within a Bagging-based ensemble.

B. *Ensemble Learning*

Ensemble learning combines multiple learning models to improve predictive performance beyond that of an individual classifier. The three most widely adopted ensemble learning techniques are Bagging, Boosting, and Stacking [8–11].

Stacking combines the predictions of multiple base models using a meta-learner, which learns how to integrate their outputs [8]. While Stacking can achieve strong predictive performance, the inclusion of a meta-learning stage introduces additional model training and computational complexity.

Bagging constructs multiple independent models using bootstrap samples generated from the original training dataset and combines their predictions through an aggregation technique such as majority voting [9]. Since each model is trained independently on a different bootstrap sample, Bagging effectively reduces prediction variance and improves model stability [9, 11].

Boosting trains models sequentially, where each successive model attempts to correct the errors of its predecessor by assigning greater importance to misclassified samples [10]. Although Boosting often improves predictive accuracy, it may be more sensitive to noisy or difficult training instances.

Among these techniques, Bagging was selected for the proposed framework because it supports independent model training, naturally generates diverse learning samples through bootstrap sampling, and helps reduce prediction variance [9, 11]. Furthermore, its independent training structure is well suited to the proposed heterogeneous ensemble, where multiple deep learning architectures are trained separately and their selected predictions are subsequently aggregated.

C. *Proposed Bagging-Based Deep Learning Framework*

The proposed methodology consists of several sequential stages designed to improve the reliability of credit card fraud detection.

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors).*

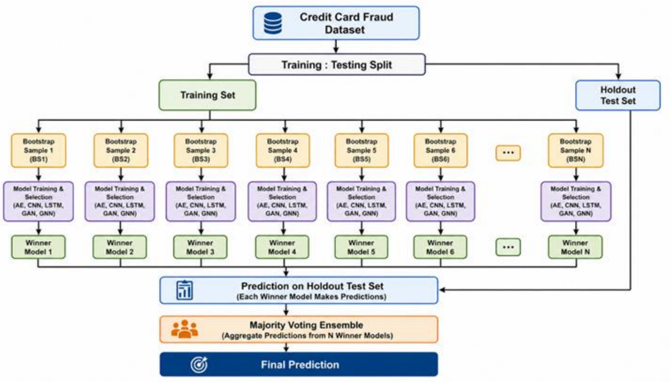

Figure 1. Proposed Bagging-Based Deep Learning Ensemble Framework

Initially, the input dataset is divided into training and independent holdout testing sets. The holdout dataset remains completely unseen during model training and is used only for the final evaluation of the ensemble.

Multiple bootstrap samples are subsequently generated from the training data using sampling with replacement. Since every bootstrap sample contains a slightly different distribution of training instances, the ensemble is able to learn diverse transaction characteristics while reducing prediction variance.

Each bootstrap sample is independently trained using five deep learning models, Autoencoder (AE), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Generative Adversarial Network (GAN), and Graph Neural Network (GNN). The trained models are evaluated using three validation strategies, 70:30 Train-Test Split, 5-Fold Cross Validation, 10-Fold Cross Validation.

Among all models trained on a bootstrap sample, the model obtaining the highest F1-score is selected as the local winner. After processing all bootstrap samples, the selected winner models are applied to the independent holdout dataset. Their predictions are finally combined through Majority Voting, producing the final fraud classification. This aggregation strategy reduces dependence on any individual classifier, minimizes prediction variance, and improves the robustness and generalization capability of the proposed fraud detection framework.

D. Deep Learning Models

1) Autoencoder (AE)

The Autoencoder is an unsupervised neural network that learns the normal characteristics of legitimate transactions by minimizing reconstruction error. Transactions exhibiting significantly higher reconstruction errors are considered anomalous and are therefore identified as potential fraudulent activities.

2) Convolutional Neural Network (CNN)

The Convolutional Neural Network extracts discriminative feature representations from transaction data through convolution operations. It automatically learns local feature interactions that contribute to distinguishing legitimate and fraudulent transactions.

3) Long Short-Term Memory (LSTM)

Long Short-Term Memory networks employ memory cells and gating mechanisms to capture long-term dependencies within transaction data. Their capability to retain relevant information enables improved learning of complex behavioral patterns associated with fraudulent activities.

4) Generative Adversarial Network (GAN)

The Generative Adversarial Network consists of a Generator and a Discriminator trained simultaneously through adversarial learning. The Generator produces synthetic transaction samples, while the Discriminator learns to distinguish genuine transactions from generated ones. Within the proposed framework, the trained Discriminator functions as the fraud detection model.

5) Graph Neural Network (GNN)

Graph Neural Networks model structural relationships among transactions by propagating information across interconnected nodes. This enables the framework to capture complex relational dependencies that may not be effectively learned by conventional deep learning architectures.

E. Winner Selection Using F1-Score

Since credit card fraud detection involves highly imbalanced datasets, model selection based solely on classification accuracy may produce misleading results. Therefore, the proposed framework employs the F1-score as the primary criterion for selecting the winner model from each bootstrap sample.

The evaluation metrics are calculated as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where TP,TN,FP and FN are True Positive, True Negative, False Positive and False Negative respectively.

For each bootstrap sample, the model achieving the highest F1-score is selected as the winner and included in the final ensemble.

F. Algorithm of the Proposed Framework

Algorithm: Bagging-Based Deep Learning Ensemble Framework

Input: Credit Card Transaction Dataset

Output: Fraud Prediction

Begin

1. Load the credit card transaction dataset.
2. Perform data preprocessing and feature scaling.
3. Split the dataset into training (80%) and holdout testing (20%) sets.
4. Generate multiple bootstrap samples from the training dataset.
5. Train AE, CNN, LSTM, GAN, and GNN independently on each bootstrap sample.
6. Evaluate all models using 70:30 split, 5-fold CV, and 10-fold CV.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

7. Compute Precision, Recall, and F1-score for each model.
8. Select the highest F1-score model from each bootstrap sample.
9. Apply all winner models to the holdout testing dataset.
10. Combine predictions using Majority Voting.
11. Generate the final fraud prediction.

End

IV. RESULTS AND DISCUSSION

A. *Experimental Evaluation*

The proposed Bagging-Based Deep Learning Ensemble Framework was evaluated on two publicly available credit card fraud datasets representing balanced and highly imbalanced transaction distributions. To comprehensively assess the robustness and generalization capability of the proposed framework, three validation strategies were employed: 70:30 Train-Test Split, 5-Fold Cross Validation, and 10-Fold Cross Validation. Performance was evaluated using Accuracy, Precision, Recall, and F1-Score, which are widely accepted metrics for fraud detection, particularly for highly imbalanced datasets. The proposed framework was also compared with several recently published machine learning and deep learning approaches to evaluate its effectiveness under different experimental settings.

B. *Dataset Description*

The proposed Bagging-Based Deep Learning Ensemble Framework was evaluated using two publicly available credit card fraud datasets representing different class distributions. The first dataset is balanced, while the second is a highly imbalanced real-world dataset. Evaluating the framework on both datasets provides a comprehensive assessment of its robustness and generalization capability under different fraud detection scenarios.

1) *Balanced Credit Card Fraud Dataset*

The balanced dataset contains 568,630 credit card transactions, equally distributed between legitimate (284,315) and fraudulent (284,315) transactions. It includes 28 anonymized PCA-transformed features along with Transaction Amount, ID, and the target variable Class. The balanced distribution enables effective evaluation of the framework without class imbalance bias.

2) *Highly Imbalanced Credit Card Fraud Dataset*

The second dataset contains 284,807 transactions collected from European cardholders, including 284,315 legitimate and 492 fraudulent transactions. It comprises 28 PCA-transformed features together with Time, Amount, and Class. The highly imbalanced nature of this dataset makes it suitable for evaluating the proposed framework under realistic fraud detection conditions.

C. *Performance Evaluation on the Balanced Dataset*

Table 2 presents the ensemble performance of the proposed framework on the balanced dataset using the three validation strategies. The framework achieved consistently high Accuracy, Precision, Recall, and F1-Score across all experiments. The minimal variation among the three validation strategies indicates that the proposed ensemble maintains stable predictive performance and generalizes well to unseen transaction data.

Table 2. Performance of the proposed framework on three validation strategy for balanced dataset.

Validation Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
70:30 Split	99.945	99.954	99.935	99.945
5-Fold Cross Validation	99.938	99.951	99.926	99.938
10-Fold Cross Validation	99.945	99.954	99.935	99.945

Table 3 compares the proposed framework with existing methods on the balanced dataset using the 70:30 Train-Test Split. The proposed framework demonstrates competitive performance across all evaluation metrics, confirming the effectiveness of integrating multiple deep learning models within a Bagging-based ensemble.

Table 3. Comparison of the proposed framework with existing methods for training-testing splitting on balanced dataset

Method	Validation strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	**70:30 Split**	99.945	99.954	99.935	99.945
Ding et al. [3]	**Train-Test**	---	---	94.85	80.27
IQR-Based CCF Detector [13]	**70:30 Split**	99.98	99.98	---	---
Logistic Regression [14]	**70:30 Split**	97.00	100.00	---	---

Table 4 summarizes the comparison under 5-Fold Cross Validation. The proposed framework consistently maintains high Accuracy, Precision, Recall, and F1-Score, indicating reliable performance across multiple validation folds.

Table 4. Comparison of the proposed framework with existing methods for 5-fold cross validation on balanced dataset.

Method	Validation strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	**5-Fold CV**	99.938	99.951	99.926	99.938
Btoush et al. [7]	**5-Fold CV**	99.10	91.70	97.70	94.63
Almalki & Masud [15]	**5-Fold CV**	99.00	~99.00	~99.00	99.00
Deep Neural Network CFD [16]	**5-Fold CV**	99.47	---	---	99.82

Table 5 presents the comparison under 10-Fold Cross Validation. The results demonstrate that the proposed framework performs competitively with recently published approaches while maintaining stable performance across repeated validation experiments.

Table 5. Comparison of the proposed framework with existing methods for 10-fold cross validation on balanced dataset

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Method	Validation strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	**10-Fold CV**	99.945	99.954	99.935	99.945
Benchmarking ML [12]	**K = 3, 5, 10 Fold**	97.12	84.44	73.57	78.63
PySpark-XGBoost-CatBoost [17]	**10-Fold CV**	99.97	---	---	99.95

D. *Performance Evaluation on the Highly Imbalanced Dataset*

Table 6 presents the ensemble performance of the proposed framework on the highly imbalanced dataset. Although the class distribution is extremely skewed, the framework achieves consistently high Accuracy and Recall across all validation strategies. The comparatively lower Precision and F1-Score reflect the inherent difficulty of fraud detection on highly imbalanced datasets while still demonstrating effective identification of fraudulent transactions.

Table 6. Performance of the proposed framework on three validation strategy on imbalanced dataset.

Validation Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
70:30 Split	99.868	57.823	86.735	69.388
5-Fold Cross Validation	99.854	54.902	85.714	66.932
10-Fold Cross Validation	99.842	52.532	84.694	64.844

Table 7 compares the proposed framework with existing studies using the 70:30 Train-Test Split. The framework achieves competitive results, particularly in terms of Recall, highlighting its capability to detect fraudulent transactions under real-world data distributions.

Table 7. Comparison of the proposed framework with existing methods for training-testing splitting on imbalanced dataset

Method	Validation strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	**70:30 Split**	99.868	57.823	86.735	69.388
Ding et al. [3]	**Train-Test**	---	---	94.85	80.27
IQR-Based CCF Detector [13]	**70:30 Split**	99.98	99.98	---	---
Logistic Regression [14]	**70:30 Split**	97.00	100.00	---	---

Table 8 summarizes the comparison under 5-Fold Cross Validation. The results indicate that the proposed framework maintains consistent predictive performance while effectively reducing the impact of class imbalance through the Bagging-based ensemble strategy.

Table 8. Comparison of the proposed framework with existing methods for 5-fold cross validation on imbalanced dataset

Method	Validation strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	**5-Fold CV**	99.854	54.902	85.714	66.932
Btoush et al. [7]	**5-Fold CV**	99.10	91.70	97.70	94.63
Almalki & Masud [15]	**5-Fold CV**	99.00	~99.00	~99.00	99.00
Deep Neural Network CFD [16]	**5-Fold CV**	99.47	---	---	99.82

Table 9 presents the comparison under 10-Fold Cross Validation. The proposed framework continues to demonstrate competitive performance compared with recent state-of-the-art approaches, confirming its robustness and practical applicability for fraud detection.

Table 9. Comparison of the proposed framework with existing methods for 10-fold cross validation on balanced dataset

Method	Validation strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	**10-Fold CV**	99.842	52.532	84.694	64.844
Benchmarking ML [12]	**K = 3, 5, 10 Fold**	97.12	84.44	73.57	78.63
PySpark-XGBoost-CatBoost [17]	**10-Fold CV**	99.97	---	---	99.95

E. *Discussion*

The experimental results demonstrate that the proposed Bagging-Based Deep Learning Ensemble Framework achieves reliable and consistent performance across both balanced and highly imbalanced credit card fraud datasets. On the balanced dataset, the framework maintained Accuracy, Precision, Recall, and F1-Score above 99.9% under all three validation strategies, with only negligible variations. This consistency indicates that the proposed framework generalizes effectively and is not dependent on a particular data partition.

On the highly imbalanced dataset, the framework achieved Accuracy between 99.842% and 99.868% while maintaining Recall above 84% across all validation strategies. Although Precision (52.532%–57.823%) and F1-Score (64.844%–69.388%) were comparatively lower due to the severe class imbalance, the high Recall demonstrates the framework's capability to successfully identify the majority of fraudulent transactions, which is a key objective in fraud detection.

The comparison with existing studies further validates the effectiveness of the proposed methodology. The integration of multiple deep learning models through Bagging, together with F1-score-based winner selection and Majority Voting, enables the framework to produce stable and competitive performance while reducing prediction variance and improving robustness.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Overall, the proposed framework demonstrates strong generalization capability and reliable fraud detection performance under both balanced and real-world imbalanced conditions, highlighting its suitability for practical credit card fraud detection applications.

V. CONCLUSION

This paper proposed a Bagging-Based Deep Learning Ensemble Framework for credit card fraud detection by integrating five complementary deep learning models, namely Autoencoder (AE), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Generative Adversarial Network (GAN), and Graph Neural Network (GNN). The framework employs bagging, F1-score-based winner selection, and majority voting to improve prediction stability, reduce model variance, and enhance the generalization capability of fraud detection systems.

The proposed framework was evaluated on both balanced and highly imbalanced credit card fraud datasets using 70:30 Train-Test Split, 5-Fold Cross Validation, and 10-Fold Cross Validation. Experimental results demonstrated consistently high performance across different validation strategies. The balanced dataset achieved Accuracy, Precision, Recall, and F1-Score values above 99.9%, while the framework maintained Accuracy above 99.84% and Recall above 84% on the highly imbalanced dataset, demonstrating its effectiveness under realistic fraud detection scenarios. Comparative analysis with recently published approaches further confirmed the competitiveness and robustness of the proposed methodology.

Overall, the proposed Bagging-Based Deep Learning Ensemble Framework provides a reliable and generalized solution for credit card fraud detection by effectively combining diverse deep learning models within a bagging ensemble. Future work will focus on evaluating the framework on larger and more diverse financial datasets, incorporating advanced deep learning architectures such as Transformer-based and Graph Attention Network (GAT) models, and extending the framework for real-time fraud detection in practical financial environments.

REFERENCES

[1] Tiwari, P., Mehta, S., Sakhuja, N., et al. (2021). *Credit Card Fraud Detection Using Machine Learning: A Study.*

[2] Naik, S. A., & Pise, N. (2022). *Credit Card Fraud Detection Using Machine Learning.*

[3] Ding, L., Liu, L., Wang, Y., Shi, P., & Yu, J. (2024). "An AutoEncoder Enhanced Light Gradient Boosting Machine Method for Credit Card Fraud Detection." *PeerJ Computer Science*, 10, e2323.

[4] Farabi, S. F., Prabha, M., Alam, M., et al. (2024). *Enhancing Credit Card Fraud Detection: A Comprehensive Study of Machine Learning Algorithms and Performance Evaluation.*

[5] *Advances in Intelligent Systems and Computing* (2020). *Enhancing Credit Card Fraud Detection Using Deep Neural Network.*

[6] *ICAART* (2025). *Enhanced Credit Card Fraud Detection Using Federated Learning, LSTM Models, and the SMOTE Technique.*

[7] Btoush, E., Zhou, X., Gururajan, R., Chan, K. C., & Alsodi, O. (2025). "Achieving Excellence in Cyber Fraud Detection: A Hybrid ML+DL Ensemble Approach for Credit Cards." *Applied Sciences*, 15(3), 1081.

[8] Wolpert, D. H. (1992). "Stacked Generalization." *Neural Networks*, 5(2), 241–259.

[9] Breiman, L. (1996). "Bagging Predictors." *Machine Learning*, 24(2), 123–140.

[10] Freund, Y., & Schapire, R. E. (1996). "Experiments with a New Boosting Algorithm." *Proceedings of the International Conference on Machine Learning (ICML).*

[11] Dietterich, T. G. (2000). "Ensemble Methods in Machine Learning." *Multiple Classifier Systems.*

[12] Shah, D., & Sharma, L. K. (2025). "Benchmarking Machine Learning Techniques for Credit Card Fraud Detection." *Indian Journal of Science and Technology*, 18(21), 1681–1695.

[13] *Sensors* (2023). "An Improved CCF Detector to Handle the Problem of Class Imbalance with Outlier Normalization Using IQR Method."

[14] *Journal of Information Systems and Technology Management* (2025). "Fraudulent Credit Card Transaction Detection Using Logistic Regression."

[15] Almalki, F., & Masud, M. (2025). "Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods." *arXiv preprint*, arXiv:2505.10050.

[16] Al-Khasawneh, M. A., Faheem, M., Alsekait, D. M., Abubakar, A., & Issa, G. F. (2025). "Hybrid Neural Network Methods for the Detection of Credit Card Fraud." *Security and Privacy*, 8(1).

[17] Theodorakopoulos, L., Theodoropoulou, A., Tsimakis, A., & Halkiopoulos, C. (2025). "Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost." *Electronics*, 14(9), 1754.