International Journal of Engineering & Tech Development
ISSN: 3069-672X/ Open Access / IJETD_Y1F2A001
Apex Academia Press / https://doi.org/10.63665/ijetd-y1f2a001

Original Article

Ensemble Machine Learning Models for Predicting Credit Card Transaction Frauds in Banking Sector

Sunil Jacob Enokkaren[1], Jaya Vardhani Mamidala[2], Varun Bitkuri[3], Avinash Attipalli[4], Raghuvaran Kendyala[5], Jagan Kurma[6]

[1] *ADP, Solution Architect*
[2] *University of Central Missouri, Department of Computer Science*
[3] *Stratford University ,Software Engineer*
[4] *University of Bridgeport, Department of Computer Science*
[5] *University of Illinois at Springfield, Department of Computer Science*
[6] *Christian Brothers University, Computer Information Systems*

Abstract

Banks are known to incur substantial financial loss every year because of financial fraud in the banks. This can be mitigated through early detection, the development of a counter-strategy, and the recuperation of losses caused by such fraud. This paper presents a proposed ensemble architecture that integrates Long Short-Term Memory (LSTM) and Artificial Neural Network (ANN) to overcome the limitations of class imbalance and multi-layered patterns in transactional data during Credit Card Fraud Detection (CCFD). With the Kaggle CCFD dataset, some preprocessing methods were performed, such as balancing data using the Synthetic Minority Oversampling Technique (SMOTE) and the top features selected using the Random Forest importance, as well as normalizing the values using Min-Max scaling. The proposed ensemble model reached a true rate of 98.67, a true accuracy of 98.51, a recall of 99.89 and an F1-score of 98.34 - far outperforming the traditional classifiers of Decision Trees (DT), Logistic Regression (LR), Naive Bayes (NBs), and K- K-Nearest Neighbors (KNN). These results demonstrate the ability of the ensemble model to be effective at modeling complex non-linear relationships, minimizing misclassification, and making predictable forecasts in extremely imbalanced data sets. The results highlight that ensemble machine learning (ML) methods have the capacity to augment current fraud detection systems and provide a foundation for future research to create stronger, larger, and safer financial fraud detection systems.

Keywords

Financial Risk Management, Anomaly Detection, Fraudulent Transactions, Ensemble Machine Learning, Data Mining Techniques, Classification Algorithms, Predictive Analytics, Banking Sector Security, Credit Card Fraud Detection.

Article History

Received: 12.03.2025

Accepted: 23.03.2025

Published: 13.04.2025

1. Introduction

Banks have quickly transitioned their services to digital platforms, replacing traditional services that involve person-to-person interactions. They provide customers with access to financial services via e-commerce, online payment systems, and internet banking. Convenience, efficiency and competitiveness within the banking industry have been highly improved through these technological advancements [1]. However, the popularity of digital transactions has also become a focus for cybercriminals, who frequently attempt fraud to exploit online financial services and breach systems. [2]. Credit-card transaction fraud (CCTF) is one such threat that has proven to be a significant challenge, resulting in substantial financial losses and eroding customer confidence. This has further increased the risk of such fraud, as online transactions can be completed using only card details, rather than the physical card. The increasing rate at which credit-card transaction fraud has become a common occurrence underscores the importance of a robust detection mechanism in the banking industry. Conventional rule-based systems can also be inadequate for detecting complex and emerging trends of fraud, since they are based on a preset set of rules that can be ineffective in identifying minor or developing fraud trends [3]. This leads to urgent demands

for smart, evidence-based solutions that can keep up with the dy.7namic threats continuously without compromising accuracy and minimizing false alarms [4].

In that regard, machine learning has also become a robust option in identifying fraudulent transactions [5][6]. With the application of large-scale data on past transactions, machine learning models will be able to unravel hidden trends and separate legitimate and fraudulent operations and cause real-time warnings of suspicious actions. Although separate algorithms, such as Decision Trees, Logistic Regression, or Neural Networks, can be relatively good at performance, they may be limited in terms of generalization and robustness [7]. To address these constraints, ensemble machine learning models that involve the combination of several classifiers have been in the limelight [8][9]. The boosting, bagging, and stacking techniques will utilize the merits of various algorithms and achieve enhanced predictive accuracy, fewer misclassifications and provide more stable solutions to CCFD in the banking industry. This paper is dedicated to the issues of implementation of ensemble ML methods to progress the detection of CCFD by means of introduction of a strong, scalable, and adaptable framework that will reinforce the financial security in the digital age.

A. Aim and Contribution

The primary aim of the study is to develop an effective and powerful machine learning model capable of detecting credit card fraud that can accommodate data imbalance, extract pertinent features, and achieve a high level of accuracy and reliability in detecting fraudulent transactions. The key contributions are:

- Balanced the highly imbalanced CCFD dataset using SMOTE to improve fraud detection performance.
- Selected influential features using Random Forest importance and applied Min-Max scaling to ensure reduced dimensionality, prevent overfitting, and maintain consistent feature contribution.
- Developed ensemble model combining ANN, and LSTM with optimized hyperparameters and a weighted voting mechanism for enhanced fraud detection performance.
- Checked the model's robustness and reliability by evaluating it with F1-score, recall, precision, accuracy, and the confusion matrix.

The justification for this study lies in the urgent need to detect rare yet highly impactful credit card fraud cases, where traditional models often fail due to severe class imbalance and complex data patterns. The novelty of this work lies in the integration of RF-based feature selection, SMOTE balancing, and Min-Max scaling ensemble models, enabling it to capture non-linear fraud patterns more effectively than conventional approaches. This combination ensures improved accuracy, balanced precision and recall, and robust generalization, making the model both reliable and practical for fraud detection in the real world.

B. Structure of Paper

The following is the outline for this part of the study. In Section 2, review the literature and background of fraud detection. Section 3 covers the methodology, dataset, and preprocessing. While Section 4 offers the results, discussion, and comparisons, Section 5, "Conclusion and Future Suggestion," goes over the findings, their limitations, and possible follow-up actions.

2. Literature Review

Research the related literature on CCFD systems and methods in this section. There are now three branches of literature in this area, each dealing with a different technique: statistical methods, deep learning techniques and ML algorithms.

Taneja, Suri and Kothari (2019) found the optimal mix of classifiers and balancing strategies by comparing and contrasting them. A European bank's regular credit card dataset was used. Among the balancing procedures that they have employed are Up sampling, Down sampling, Borderline SMOTE, Regular SMOTE, ADASYN, and SVM SMOTE. For this reason, they have compared using classifiers like boosting and models. With a 0.85 F-score, the outcomes demonstrated that optimal method for balancing datasets was to use SVM SMOTE in conjunction with a RF classifier [10].

Kumar et al. (2019) used RFA, or the RF Algorithm, to detect legitimate and fraudulent purchases. This method uses decision trees to categorise datasets, and it is based on supervised learning methods. A confusion matrix is

generated following the dataset's classification. Using the confusion matrix, they may assess how well the Random Forest Algorithm performs. Processing the dataset yielded findings with an accuracy of approximately 90% [11].

Sethia, Patel and Raut (2018) improve model performance by using several adversarial networks to produce pseudo data. Methods from the GAN vanilla implementation, Margin Adaptive, Least Squares, and Relaxed Wasserstein are employed in this study. Optimal data creation frequencies, classifier accuracy, model convergence, and data dispersion compared to original fraud data are all examined. After that, the data is tweaked and evaluated with an ANN model; for an initial class imbalance dataset of 579 to 1, the recall increases by 12.86% [12].

Gyamfi and Abdulai, (2018) analyse the various forms of bank fraud and how data mining technology can aid in the early detection of these frauds by utilising the data that banks generally have on hand. We utilise supervised learning methods. To determine the legitimacy of proposed financial transactions, use SVM with Spark (SVM-S) to train models that capture normal and unusual user actions. Experiment findings demonstrate that SVM-S is more effective at making predictions than Back Propagation Networks. When the ratio of training data to total data approaches 0.8, both accuracy and average prediction reach their maximum potential [13].

Zamini and Montazer (2018) proposed a procedure for unsupervised fraud detection that utilizes autoencoder-driven clustering. A three-hidden-layer autoencoder with k-means clustering was applied on 28,4807 transactions originating from European banks. When compared to other methods, this one performed better, with a 98.9% accuracy rate and an 81% TPR [14].

The reviewed studies (Table 1) demonstrate progress in CCFD using statistical, ML, and DL approaches, yet key gaps persist. Most models struggle with extreme class imbalance, scalability on real-time banking data, and lack of interpretability. While balancing methods and ensemble models improve performance, they add computational overhead, and DL models like GANs and autoencoders face high training complexity. Thus, there remains a need for hybrid, scalable, and explainable frameworks that can effectively detect rare fraud patterns.

Table 1: Summary of Existing Work on CCFD using ML

Author	Technique / Data	Key Findings	Challenges	Recommendations
Taneja, Suri & Kothari (2019)	Balancing techniques (Down/Up Sampling, Borderline SMOTE, SMOTE, SVM SMOTE, ADASYN) + Bagging/Boosting on the European credit card dataset	SVM SMOTE + Random Forest achieved best performance (F-score = 0.85)	Severe class imbalance; computational overhead with multiple resampling techniques	Combine advanced balancing (e.g., SVM SMOTE) with ensemble models for better fraud detection
Kumar et al. (2019)	Random Forest Algorithm on the credit card dataset	Achieved ~90% accuracy in detecting fraudulent transactions	Accuracy alone not sufficient; lacks deeper analysis with precision/recall	Incorporate multiple metrics (precision, recall, F1) and compare with other ensemble classifiers
Sethia, Patel & Raut (2018)	Multiple GANs (Vanilla, LSGAN, WGAN, MAGAN, RWGAN) + ANN classifier on highly imbalanced dataset (579:1 ratio)	GAN-generated pseudo-data improved recall by 12.86%	Training instability of GANs requires extensive computational resources	Employ GANs for data augmentation in extreme imbalance cases; hybrid with ANN/DL models
Gyamfi & Abdulai (2018)	Support Vector Machines with Spark (SVM-S) vs. Backpropagation Networks on bank transaction datasets	SVM-S outperformed BPN; maximum accuracy at 0.8 train ratio	Limited scalability with extremely large datasets; Spark implementation complexity	Use distributed frameworks (Spark, Hadoop) for real-time fraud detection at scale

Zamini & Montazer (2018)	Autoencoder (3 hidden layers) + k-means clustering on European dataset (284,807 records)	Achieved 98.9% accuracy and 81% TPR	High accuracy but lacks interpretability; clustering sensitive to initialization	Combine deep autoencoders with interpretable ML for trustworthy fraud detection

3. Methodology

The methodology begins with using the Kaggle CCFD dataset, followed by pre-processing steps such as handling missing/null values, balancing the highly imbalanced data with SMOTE, and applying RF feature importance to select top 27 features. Min-Max scaling is then utilised to regularize the features, and the dataset is divided into 20% testing and 80% training sets. An ensemble model combined with ANN and LSTM is employed for fraud detection. Finally, utilise accuracy, precision, recall, F1-score, and confusion matrix analysis to assess the model's performance and ensure the reliability of fraud detection.

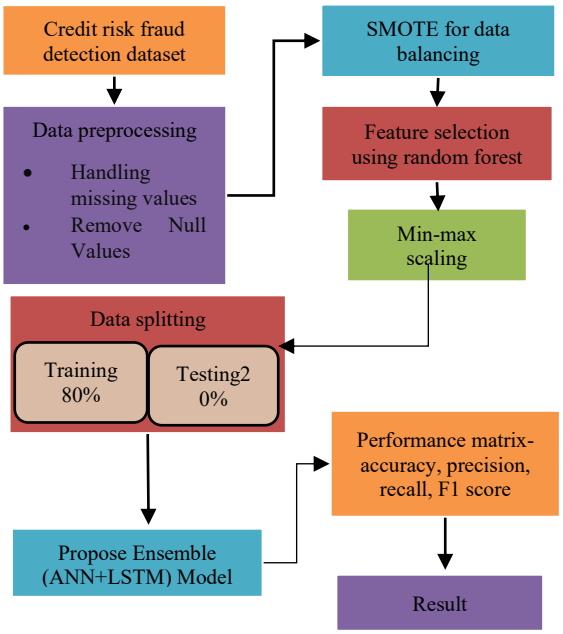

Fig-1: Propose Flowchart for Credit Card Fraud Detection

Each step of implementation in Figure 1 is discussed below;

A. Data Collection

Kaggle CCFD provided the dataset that they utilised for this fraud detection investigation. The positive class is severely underrepresented in the dataset, comprising 284,807 transactions, with 492 being fraudulent (0.172%). There are thirty features in the dataset. These include Time and Amount in addition to twenty-eight anonymised principal components (V1-V28) obtained via principal component analysis.

B. Data Pre-Processing

Cleaned and transformed data were used in many pre-treatment stages before machine learning methods were applied to address the issue of detecting fraud. Here are the steps:

- Handling missing values: If any records in the dataset are missing values, you can either remove them or impute them.
- Remove Null Values: The characteristics is_ftp_login, attack_cat, and ct_flw_http_mthd have had their null values removed to ensure the study's accuracy.

C. Data Balancing with SMOTE

The minority class (non-fraud) is subjected to the extremely successful SMOTE to rectify the class imbalance [15]. For the minority group, SMOTE primarily aims to generate synthetic data points while preserving the inherent patterns and correlations of the original data.

Fig-2: Before and After SMOTE Data Distribution

The dataset's class distribution both before and after SMOTE application is shown in Figure 2. It is challenging to detect fraud at the outset, as the dataset is skewed towards legitimate transactions (284,315) rather than fraudulent ones (492). There are 492 instances of fraud and 492 cases of non-fraud after using SMOTE, since the minority class is oversampled to coincide with class that is in majority. This balanced dataset makes it easier to train ML models fairly, which in turn increases their capacity to identify unusual fraudulent behaviours.

D. Feature Selection with Random Forest

The initial run uses Random Forest feature significance to choose the top 27 features. The goals of this choice are to save training time, prevent overfitting, and enhance the model's predictive power. Using the Random Forest significance metric, the top 27 features are displayed in Figure 3.

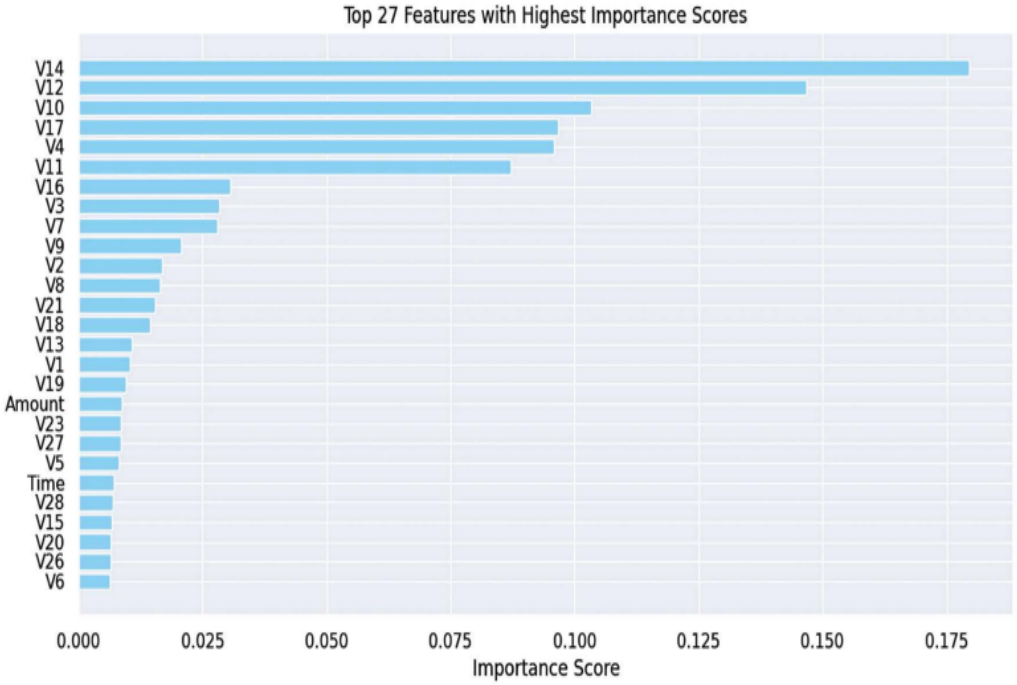

Fig-3: Top 27 Features' Importance Score

Figure 3 displays top 27 features ranked by their value scores in a predictive model. The features are primarily labeled as V1 to V28, most probably developed from methods for reducing dimensionality, along with original features such as "Amount" and "Time." Among them, V14, V10, and V17 stand out with the highest importance scores, indicating their strong influence on model's decision-making. This visualization helps identify which characteristics have the most impact on the model's accuracy, aiding in feature selection and model interpretability.

E. Min-Max Scaling

In this work, characteristics were adjusted to a scale from 0 to 1 using Min-Max normalization approaches. Equation (1) can be used to express normalization mathematically:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (1)$$

where x is the starting point, x' is the modified value, niZ (x) is the dataset's minimum value, and max (x) is the feature's maximum value.

F. Data Splitting

There is an 80/20 division in the dataset, with 20% used for training and 20% for testing. Instead of using the test data to build the model, it is used to verify the model's performance after training. The models are constructed and trained using the training data.

G. Propose Ensemble (ANN+LSTM) Model

The planned ensemble model is a fusion of LSTM and ANN networks that can be used to enhance predictive performance by combining the high-quality nonlinear feature learning of the ANN with the sequential pattern learning of the LSTM networks. In this hybrid architecture, the ANN uses dense layers with Re LU activation to manipulate the static or tabular features, and LSTM operates to uncover the temporal relationships between sequential inputs; the results of both are combined and fed to a final dense prediction layer.

The Adam optimizer is used in model training by early stopping to avoid overfitting and hyperparameter tuning is performed by grid or Bayesian optimization. An ANN's primary settings include its hidden layer count (2-4), neuronal count per layer (64-256), learning rate (1e-3 to 1e-4), and dropout rate (0.2-0.5). In the case of LSTM, tuning of the number of units (32-128), sequence length, dropout rate and batch size (32-128) are done to achieve a balance between generalization and convergence rate. Additional ensemble-level fusion parameters, including the weight contribution of ANN output to LSTM output, are optimized to obtain the best trade-off between sequential and non-sequential learning. This method of systematic training and hyperparameter optimization ensures that the ANN+LSTM ensemble effectively captures both stationary and dynamic patterns, thereby outperforming standalone models.

H. Performance Matrics

The confusion matrix provides important metrics for evaluating classification models, like F1-score, recall, precision, and the accuracy. With the use case of CCFD in mind, the following can be defined using four important terms taken from a confusion matrix:

- True Positives (TP): Predicted positive occurrences with high accuracy (e.g., correctly identifying fraudulent transactions as fraud).
- True Negatives (TN): Identified valid transactions as legitimate, for example, as an example of a bad instance that was accurately predicted.
- False Positives (FP): False positive predictions (Type I error), for example, real transactions mistakenly marked as fraudulent.
- False Negatives (FN): False negative predictions (Type II mistake, for example, when the model fails to detect fraudulent transactions).

Then, the performance metrics can be defined as Eequations (2) to (5),

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$F1-score = 2 * \frac{(precision*recall}{(precision+recall)} \quad (5)$$

Accuracy is how much the model is generally correct, whereas precision is concerned with minimizing false positives- which is important to prevent false alarms in intrusion detection. Recall tests the capability to identify real fraud and this information assists in reducing the likelihood of detection. Precision and recall are normalized using the F1-score, which indicates the general model performance.

4. Result And Discussion

Hardware and cloud resources were used to support the experimental setup. The local PC was fitted with an Intel duel Core with i7 processor with a speed of 2.50 GHz that had RAM of 16 GB and would be effective in carrying out the required tasks. Table II shows the performance of the suggested Ensemble model of CCFD and indicates the effectiveness of the model in major evaluation parameters. The model achieved remarkable accuracy of 98.67%, indicating the overall reliability of the model in identifying legitimate and fraudulent transactions. The model minimizes false positives by 98.51, and the recall of 99.89 percent is a good value to make sure that almost all fraud cases are discovered. F1-score value of 98.34 percent describes an appropriate ratio of recall to accuracy, which proves that the Ensemble model has strong applicability to real-world fraud detection.

Table 2: Propose Model Performance For CCFD

Performance Matrix	Ensemble (ANN+LSTM)
Accuracy	98.67
Precision	98.51
Recall	99.89
F1-Score	98.34

Fig-4: Training and Validation Loss Curves for Ensemble

Figure 4 illustrates validation and training loss curves of Ensemble model in 50 epochs. The two curves exhibit a steady pattern of decreasing, which is evidence of successful learning and model convergence. Initially, the training loss reduces to a minimum, while the validation loss is more erratic due to the complex nature of the data. However, the total losses stabilize at approximately 0.04. The proximity of the two curves indicates that Ensemble model has a high generalization without remarkable overfitting.

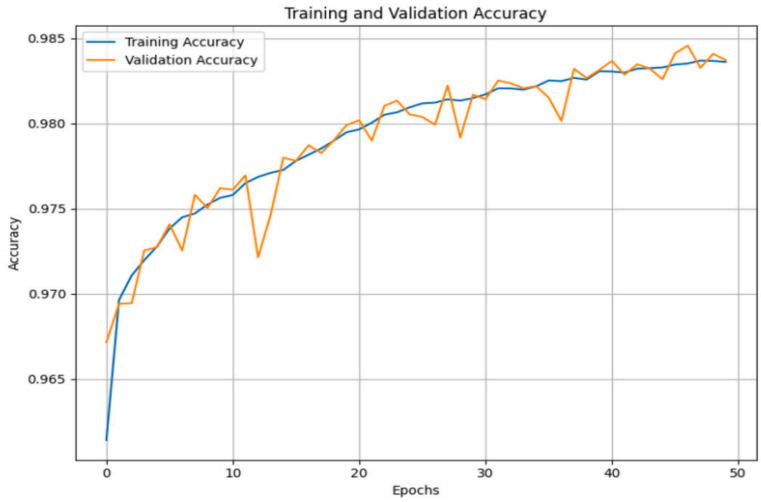

Fig-5: Training and Validation Accuracy for Ensemble Model

The accuracy curves for validation and training of the Ensemble model during 50 epochs are displayed in Figure 5. The trend in both curves is increasing steadily showing that the model's performance improves steadily over time. Validation accuracy closely mirrors training accuracy, with only slight variations, reflecting the model's robustness and generalizing ability well. By final epochs, the accuracy stabilizes around 98.5%, demonstrating the effectiveness of the Ensemble in detecting patterns within the dataset.

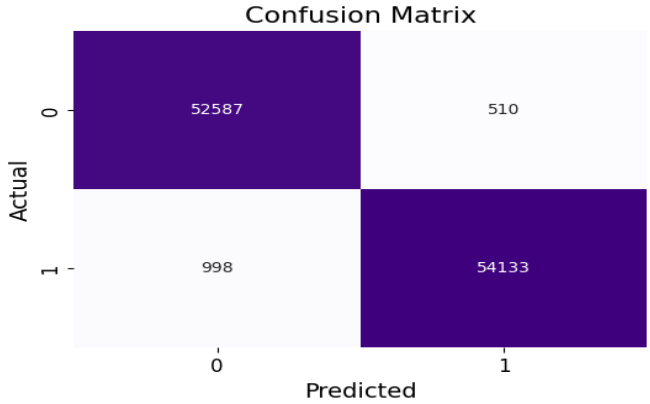

Fig-6: Confusion Matrix for Ensemble Model

Figure 6 illustrates confusion matrix for the Ensemble model, showing strong classification performance with a high amount of correctly predicted instances in both classes. The model accurately classified 52,587 true negatives and 54,133 true positives, while misclassifying only 510 false positives and 998 false negatives. This balance indicates that the Ensemble achieved reliable detection capability with minimal misclassification, demonstrating robustness in distinguishing between the two classes.

A. Comparative Analysis

The results of comparing several models for CCFD are presented in Table 3, showing notable variations in accuracy. Logistic Regression (54.86%) performed the weakest, indicating its limitations in handling imbalanced fraud data. Decision Tree (95.5%) and KNN (94.2%) delivered better results but still fell short compared to advanced methods. Naïve Bayes (97.70%) showed strong performance, though slightly lower than the Ensemble with 98.67%, which outperformed all models. This highlights the Ensemble model's superior ccapability to capture complicated fraud trends, making it the most effective among the compared approaches.

Table 3: Comparative Analysis for CCFD

Model	Accuracy
Decision Tree [15]	95.5
Logistic Regression [16]	54.86
Naïve Bayes [17]	97.70
K-Nearest Neighbors [18]	94.2
Propose Ensemble	98.67

The proposed Ensemble model proves to be quite efficient in CCFD, as it is more accurate and robust than traditional models. It can detect complex, non-linear fraud patterns, which have minimized false positives and prevented cases that would have been missed; hence, it is very reliable in real-life applications. The major strengths are high overall performance compared to baseline models, high generalization without overfitting, and the balanced precision-recall trade-off all of which contribute to the increased credibility and effectiveness in fraudulent transaction identification.

5. Conclusion

There has been an increase in attacks by fraudsters on credit card transactions compared to the past. The further development of data science and machine learning has enabled the creation of numerous algorithms to identify fraudulent transactions. In this paper, an ensemble-based method for CCFD is described, which showed impressive results in identifying fraudulent transactions with a 98.67% success rate. The model effectively struck a balance

between accuracy and recognition and thus minimized FP and FN, which is paramount to real-life uses where a false miss or false alarm might lead to a loss of money or customer dissatisfaction. Compared to traditional models like DT, LR, KNN, and NBs, the Ensemble demonstrated superior performance by capturing complex, non-linear fraud patterns, thereby proving its robustness and suitability for real-world detection. However, the research is limited by its reliance on a single dataset and synthetic oversampling with SMOTE, which may not fully reflect real-world scenarios. Future work will focus on testing with larger, more diverse datasets, exploring hybrid models such as CNN-LSTM for improved feature learning, and applying federated learning to enhance scalability, privacy, and adaptability.

6. References

[1] Z. M. Sanusi, M. N. F. Rameli, and Y. M. Isa, "Fraud Schemes in the Banking Institutions: Prevention Measures to Avoid Severe Financial Loss," Procedia Econ. Financ., 2015, doi: 10.1016/s2212-5671(15)01088-6.

[2] Y.-J. Chen, W.-C., Liou, Y.-M. Chen, and J.-H. Wu, "Fraud detection for financial statements of business groups," Int. J. Account. Inf. Syst., vol. 32, pp. 1–23, Mar. 2019, doi: 10.1016/j.accinf.2018.11.004.

[3] F. Carcillo, Y. A. Le Borgne, O. Caelen, and G. Bontempi, "Streaming active learning strategies for real-life credit card fraud detection: assessment and visualization," Int. J. Data Sci. Anal., 2018, doi: 10.1007/s41060-018-0116-z.

[4] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," Expert Syst. Appl., vol. 51, pp. 134–142, Jun. 2016, doi: 10.1016/j.eswa.2015.12.030.

[5] M. S. P, A. Saini, S. Ahmed, and S. D. Sarkar, "Credit Card Fraud Detection using Machine Learning and Data Science," Int. J. Eng. Res., vol. 08, no. 09, Sep. 2019, doi: 10.17577/IJERTV8IS090031.

[6] S. V. Suryanarayana, B. Gn, and G. V. Rao, "Machine Learning Approaches for Credit Card Fraud Detection," Int. J. Eng. Technol., vol. 7, no. 2, p. 917, Jun. 2018, doi: 10.14419/ijet.v7i2.9356.

[7] S. Carta, G. Fenu, D. R. Recupero, and R. Saia, "Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model," J. Inf. Secur. Appl., vol. 46, pp. 13–22, Jun. 2019, doi: 10.1016/j.jisa.2019.02.007.

[8] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 8, pp. 3784–3797, Aug. 2018, doi: 10.1109/TNNLS.2017.2736643.

[9] I. González-Carrasco, J. L. Jiménez-Márquez, J. L. López-Cuadrado, and B. Ruiz-Mezcua, "Automatic detection of relationships between banking operations using machine learning," Inf. Sci. (Ny)., 2019, doi: 10.1016/j.ins.2019.02.030.

[10] S. Taneja, B. Suri, and C. Kothari, "Application of Balancing Techniques with Ensemble Approach for Credit Card Fraud Detection," in 2019 International Conference on Computing, Power and Communication Technologies (GUCON), 2019, pp. 753–758.

[11] M. S. Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika, and E. Aswini, "Credit Card Fraud Detection Using Random Forest Algorithm," in 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), IEEE, Feb. 2019, pp. 149–153. doi: 10.1109/ICCCT2.2019.8824930.

[12] A. Sethia, R. Patel, and P. Raut, "Data Augmentation using Generative models for Credit Card Fraud Detection," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), IEEE, Dec. 2018, pp. 1–6. doi: 10.1109/CCAA.2018.8777628.

[13] N. K. Gyamfi and J.-D. Abdulai, "Bank Fraud Detection Using Support Vector Machine," in 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, Nov. 2018, pp. 37–41. doi: 10.1109/IEMCON.2018.8614994.

[14] M. Zamini and G. Montazer, "Credit Card Fraud Detection using autoencoder-based clustering," in 9th International Symposium on Telecommunication: With Emphasis on Information and Communication Technology, IST 2018, 2018. doi: 10.1109/ISTEL.2018.8661129.

[15] N. Khare and S. Y. Sait, "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models," Int. J. Pure Appl. Math., vol. 118, no. 20, pp. 825–838, 2018.

[16] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in Proceedings of the IEEE International Conference on Computing, Networking and Informatics, ICCNI 2017, 2017. doi: 10.1109/ICCNI.2017.8123782.

[17] 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, 2018, doi: 10.1109/ACCESS.2018.2806420.

[18] S. Shirgave, C. J. Awati, R. More, and R. More, "A review on credit card fraud detection using machine learning," Int. J. Sci. Technol. Res., vol. 8, no. 10, pp. 1217–1220, 2019.

[19] Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., Kakani, A. B., & Nandiraju, S. K. K. (2024). A Machine Learning-Based Framework for Predicting and Improving Student Outcomes Using Big Educational Data (Approved by ICITET 2024 Conference Proceedings). Available at SSRN 5315635.

[20] Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. (2025). Towards Early Forecast of Diabetes Mellitus via Machine Learning Systems in Healthcare. European Journal of Technology, 9(1), 35-50.

[21] Chalasani, R., Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Tyagadurgam, M. S. V. (2025). Big Data-Driven Approach for Lung Cancer Identification via Advanced Deep Transfer Learning Models. European Journal of Technology, 9(1), 51-67.

[22] Vattikonda, N., Gupta, A. K., Polu, A. R., Narra, B., Buddula, D. V. K. R., & Patchipulusu, H. H. S. (2024). Machine Learning-Based Approaches for Detecting and Mitigating Distributed Denial of Service (DDoS) Attacks to Improved Cloud Security. European Journal of Technology, 8(6), 28-48.

[23] Polu, A. R., Narra, B., Buddula, D. V. K. R., Hara, H., Patchipulusu, S., Vattikonda, N., & Gupta, A. K. Analyzing The Role of Analytics in Insurance Risk Management: A Systematic Review of Process Improvement and Business Agility.

[24] Madhura, R., Varshitha, P., Nikitha, S., Niveditha, K. M., & Bhat, M. (2024, December). RTL design of 16-bit RISC Processor Using Vedic Mathematics. In 2024 IEEE 33rd Asian Test Symposium (ATS) (pp. 1-4). IEEE.

[25] Harinandan, R., Kumar, M., Vamshi, P., Padma, C. R., Krishnappa, K. H., & Raghunandan, J. R. (2024, August). Design and Development of a Real-time Monitoring System for ACL Injury Prevention. In 2024 2nd International Conference on Networking, Embedded and Wireless Systems (ICNEWS) (pp. 1-6). IEEE.

[26] Krishnappa, K. H. (2024). Traffic pattern analysis for malicious node detection in NoC design. Journal of Communications, 9, 12.

[27] Mukund Sai Vikram Tyagadurgam, Venkataswamy Naidu Gangineni, Sriram Pabbineedi, Mitra Penmetsa, Jayakeshav Reddy Bhumireddy, et al. (2024) AI-Powered Cybersecurity Risk Scoring for Financial Institutions Using Machine Learning Techniques. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-482. DOI: doi.org/10.47363/JAICC/2024(3)452

[28] Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Vangala, S. R., Polam, R. M., & Kamarthapu, B. (2025). Adversarial Machine Learning in Cybersecurity: A Review on Defending Against AI-Driven Attacks. European Journal of Applied Science, Engineering and Technology, 3(4), 4-14.

[29] Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2025). Using Artificial Intelligence-Based Machine Learning Regression Models for Predictions of Home Prices. European Journal of Applied Science, Engineering and Technology, 3(3), 404-416.

[30] Nandiraju, S. K. K., Chundru, S. K., Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., & Kakani, A. B. (2025). Enhancing Cybersecurity: Zero-Day Attack Detection in Network Traffic with Deep Learning Model. Asian Journal of Research in Computer Science, 18(7), 262-273.

[31] Polam, R. M., Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., & Vangala, S. R. (2025). Advanced Machine Learning for Robust Botnet Attack Detection in Evolving Threat Landscapes. Asian Journal of Research in Computer Science, 18(8), 1-14.

[32] Kamarthapu, B., Penmetsa, M., Reddy, J., Chalasani, R., Vangala, S. R., & Polam, R. M. Data-Driven Detection of Network Threats using Advanced Machine Learning Techniques for Cybersecurity.

[33] Chundru, S. K., Vikram, M. S., Naidu, V., Pabbineedi, S., Kakani, A. B., & Nandiraju, S. K. K. Analyzing and Predicting Anaemia with Advanced Machine Learning Techniques with Comparative Analysis.

[34] Gangineni, V. N., Tyagadurgam, M. S. V., Pabbineedi, S., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2025). Preventing Phishing Attacks Using Advanced Deep Learning Techniques for Cyber Threat Mitigation. Journal of Data Analysis and Information Processing, 13(03), 10-4236.

[35] Kalla, D., Mohammed, A. S., Boddapati, V. N., Jiwani, N., & Kiruthiga, T. (2024, November). Investigating the Impact of Heuristic Algorithms on Cyberthreat Detection. In 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT) (Vol. 1, pp. 450-455). IEEE.

[36] Gangineni, V. N., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., & Pabbineedi, S. (2025). Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce. Available at SSRN 5478047.

[37] Polu, A. R., Narra, B., Buddula, D. V. K. R., Patchipulusu, H. H. S., Vattikonda, N., & Gupta, A. K. (2025). The Role of the Internet of Things in Smart Cities: Current Implementations and Pathways for Future Development. Universal Library of Engineering Technology, 2(2).

[38] Narra, B., Gupta, A. K., Buddula, D. V. K. R., Patchipulusu, H. H. S., Vattikonda, N., & Polu, A. R. (2025). Applications of Blockchain in Software Engineering: Enhancing Security, Traceability, and Transparency. International Journal of Innovative Computer Science and IT Research, 1(02), 63-75.

[39] Vattikonda, N., Gupta, A. K., Polu, A. R., Narra, B., Buddula, D. V. K. R., & Patchipulusu, H. H. S. (2025). Leveraging Deep Learning for Personalized Fashion Recommendations Using Fashion MNIST. International Journal of Emerging Trends in Computer Science and Information Technology, 6(2), 36-46.

[40] Buddula, D. V. K. R., Patchipulusu, H. H. S., Vattikonda, N., Gupta, A. K., Polu, A. R., & Narra, B. (2025). Machine Learning-Based Detection and Prevention of Anti-Money Laundering (AML) in the Financial Sector. International Journal of Innovative Computer Science and IT Research, 1(02), 53-63.

[41] Polu, A. R., Narra, B., Vattikonda, N., Gupta, A. K., Buddula, D. V. K. R., & Patchipulusu, H. H. S. AI-POWERED SYNTHETIC COGNITION NETWORKS Leveraging Multi-Agent Machine Learning to Simulate and Optimize Human Decision-Making in Complex Crisis Scenarios. Global Pen Press UK.

[42] Mitra Penmetsa, Jayakeshav Reddy Bhumireddy, Rajiv Chalasani, Mukund Sai Vikram Tyagadurgam, Venkataswamy Naidu Gangineni, Sriram Pabbineedi. (2025) Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce. International Journal of Computers, 10, 260-267

[43] Penmetsa, M., Bhumireddy, J.R., Chalasani, R., Vangala, S.R., Polam, R.M. and Kamarthapu, B. (2025) Effectiveness of Deep Learning Algorithms in Phishing Attack Detection for Cybersecurity Frameworks. Journal of Data Analysis and Information Processing, 13, 331-346. https://doi.org/10.4236/jdaip.2025.133021

[44] Prabakar, D., Iskandarova, N., Iskandarova, N., Kalla, D., Kulimova, K., & Parmar, D. (2025, May). Dynamic Resource Allocation in Cloud Computing Environments Using Hybrid Swarm Intelligence Algorithms. In 2025 International Conference on Networks and Cryptology (NETCRYPT) (pp. 882-886). IEEE.

[45] Nagaraju, S., Johri, P., Putta, P., Kalla, D., Polvanov, S., & Patel, N. V. (2025, May). Smart Routing in Urban Wireless Ad Hoc Networks Using Graph Attention Network-Based Decision Models. In 2025 International Conference on Networks and Cryptology (NETCRYPT) (pp. 212-216). IEEE.

[46] NR, A. R., Rajasri, T., Praveen, R., Kalla, D., Bendale, S. P., & Venu, N. (2025, April). CAC Training-A Unified Cybersecurity Training Program for Military Staff. In 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI) (Vol. 3, pp. 569-573). IEEE.

[47] Kalla, D., Smith, N., & Samaah, F. (2025). Deep Learning-Based Sentiment Analysis: Enhancing IMDb Review Classification with LSTM Models. Available at SSRN 5103558.

[48] Sreeramulu, M. D., Mohammed, A. S., Kalla, D., Boddapati, N., & Natarajan, Y. (2024, September). AI-driven Dynamic Workload Balancing for Real-time Applications on Cloud Infrastructure. In 2024 7th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 7, pp. 1660-1665). IEEE.

[49] Kalla, D., & Samaah, F. (2023). Exploring Artificial Intelligence and Data-Driven Techniques for Anomaly Detection in Cloud Security. Available at SSRN 5045491.