Journal of Global Research in Electronics and Communication
Volume 2, No. 1, January 2026
Available Online at: www.jgrec.info
RESEARCH PAPER

Comprehensive Study of Machine Learning Approach for Fraudulent Identification in Real-Time Financial Banking Systems

Sandeep Gupta
SATI, Vidisha
Sandeepguptabashu@gmail.com

***Abstract*—Financial fraud, which is commonly defined as applying fraudulent techniques to secure funds, has, in recent years, become a significant issue for businesses and organizations. Efforts to weed out such scams by existing means like human inspection and checks are tedious, expensive and subject to inaccuracies. In the near future, advances in artificial intelligence may enable more sophisticated machine learning algorithms to search through massive amounts of financial data for signs of fraud. This study suggests a thorough method for identifying financial fraud using the highly unequal class-marked IEEE CIS Fraud Detection dataset. A Light Gradient Boosting Machine (LightGBM) and a Convolutional Neural Network (CNN) are the two models that are subsequently trained and assessed on a balanced dataset. The results indicate that CNN model is reliable and significantly higher than LightGBM model, CNN model is 99.73% and its accuracy is higher at 99.91%. Moreover, the CNN model has low false negatives and zero false positives in the confusion matrix, which shows its usefulness in the classification of fraud transactions. The results confirm the effectiveness of CNNs to perform this operation, which gives a solid and significantly effective solution to the problem of real-time financial fraud detection.**

Keywords—Financial Identification, Fraud Detection, Machine Learning (ML), Real-Time Banking Systems, Deep Learning (DL), IEEE-CIS Dataset.

I. Introduction

Banks and other financial institutions have long been one of the primary impetuses of global economic growth by implementing technological innovation. Financial institutions have been the main source of credit. The financial industry is dominated by banks as they offer monetary and financial services [1], [2], [3]. Banking industry has been making steady investments in technology, thus making it become a significant force behind the increase in technology. In the past decade, fintech companies have made banks accelerate their digital transformation, thus adding competition and promoting spillover of innovation [4], [5]. Nonetheless, in tandem with the growth of digital banking facilities, financial fraud has become a ubiquitous and more advanced danger. Such types of frauds like identity theft [6], [7], transaction fraud, money laundering and phishing attacks are all types of fraud activities that lead to huge financial losses both to banks and consumers, as well as damage the integrity and trustworthiness of the financial system [8], [9]. Simple fraud detection and prevention is now a compelling problem because of the challenges brought about by real-time banking systems where fraudsters are operating and executing their activities in the millions, necessitating a quick and precise detection and prevention of fraud [10], [11], [12].

Financial fraud refers to the act of coaxing someone to separate with his or her money [13]. The banking, insurance, tax, and commercial are the potential victims of financial fraud. In recent years, there have been growing difficulties in fighting financial crimes in several sectors and businesses, including money laundering and other financial transactions that are characterized by fraud [14], [15]. Although there are a number of efforts to reduce financial fraud, the persistence of this issue affects the economy and society negatively as it consumes a significant amount of money on a daily basis [16], [17]. Machine learning has swept the world of fraud detection by allowing algorithms to identify suspicious patterns of both historical and real-time transaction data [18]. A more advanced implementation of this ML, Deep Learning (DL), goes a step further and can learn more subtle and non-linear relationships in fraud behavior [19], [20], [21]. Despite the remarkable achievements in the area of identifying complex malicious actions, DL is limited by computer requirements and interpretability [22]. To evaluate the benefits and drawbacks of both machine learning and deep learning techniques in relation to real-time fraud detection in banking systems, the given comparative study illuminates the possibilities of these techniques application and the effective performance of fraud combating on a great scale.

A. Motivation and Contribution of the Study

The frequency of financial crime has been making modern real-time banking systems more susceptible to fraud by increasing the complexity of the schemes targeting their customers and financial institutions. This is what has influenced this research. The old rule-based detection techniques are not able to keep pace with the altering fraud trends and are often likely to miss on the positives and the negatives. The increase in digital transactions and the abundance of transactional data is driving an urgent academic and industrial demand of detecting fraud using complex, data-driven approaches. By means of better prediction rates, minimization of false alarms, and real-time flexibility, the study presents the opportunity to implement the most advanced ML and DL solutions, in order to better secure against fraud. The suggested research helps to develop a scalable and interpretable framework that allows for conducting protected, real-time operations in the banking sector. The study's primary contributions are as follows:

- Developed a workable system for using the dataset of IEEE-CIS Fraud Detection to the detection of financial fraud.

© JGREC 2026, All Rights Reserved

- This study provides a clear methodology for handling raw financial data, including feature removal, median imputation, and min-max normalization..
- It highlights the successful application of the SMOTE technique to resolve the fraud detection dataset's stark class disparity.
- Developed two advanced models, LightGBM and CNN, for real-time fraud detection.
- To assess models using f1-score, ROC, recall, accuracy, and precision.

B. Structure of the Paper

The structure of the paper as follows: Section II covers the current literature on financial fraud detection, Section III details the methodology, Section IV, result analysis and a comparative analysis of the models is discussed, Lastly, Section V concludes the research and identifies potential areas for future work.

II. Literature Review

This section provides an overview of previous research on financial fraud detection in real-time banking systems. Numerous algorithmic techniques have been employed by various researchers in the literature with the goal of enhancing the diagnostic process's speed, accuracy, and dependability. Some of the main themes that are identified in these studies are:

Mishra, Biswal and Padhy (2025) detect financial system fraud using a variety of machine learning classifiers. Decision trees, AdaBoost, Gradient Boosting, SVM, KNN, RF, and LR were among the classifiers utilized. F1-score, recall, accuracy, and precision are these classifier metrics. They concluded that KNN, RF, and recall (98.8937, 98.5, and 98.5000) were the highest scoring classifier, the highest accuracy and the highest recall respectively of their experiment. It suggested that RF is the most suitable among the other classifiers to detect fraud in the bank system. At the same time, the AdaBoost and Gradient Boosting classifiers have a good precision and AUC-ROC values [23].

Souran and Shah (2025) introduce a new hybrid ML model that improves the accuracy of detection of fraud and reduces the number of false positives. Supervised learning (XGBoost) and unsupervised anomaly detection (Isolation Forest) are combined in the model. The model well outperformed the DL and traditional ML models with an AUC-ROC of 0.996 percent and an accuracy of 99.1 percent calculated using a publicly available financial transaction dataset. The hybrid strategy takes advantage of both behavioral and transactional qualities, which offers flexibility to the changing fraud patterns [24].

R et al. (2025) presented a Stacking Ensemble Model (SEM) to effectively detect financial fraud, which takes into account the diversity of base models to handle drift patterns. Initially, fraudulent data is collected from CCFD dataset, which is openly accessible at Kaggle. The next step in pre-processing is to use sampling techniques to ensure that the input data is balanced. Finally, the proposed SEM model is incorporated to detect financial fraud by fusing the predictions from diverse base models. From the results, the proposed SEM model outperformed existing Boosting Techniques in terms of accuracy (99.8%), AUC (93.5%), respectively [25].

Keerthana et al. (2024) introduced a DL-based model for digital banking fraud detection that uses a plethora of CNN and a succession of RNNs to learn patterns from real-time transaction data. The model generates highly accurate values of 95.2%, as well as precisions and recalls of 91.0% and 89.3%, respectively. Furthermore, the model generates better results in comparison to the systems used in existing applications, which produce an average accuracy of about %. The processing time of the model is relatively short, which takes about 15ms, and in the meantime, the model generates low levels of false positives and relatively high rates of real fraud detection, which are assessed at 65% [26].

Narejo et al. (2024) designed architecture of the model is based on LSTM and Sequence Time Series Transformer (SeqTrans). Firstly, aimed to remove outliers and dataset imbalance, thereby refining the quality of the datasets for model training. Secondly, the proposed model, LSTM and SeqTrans, helps to eliminate fraud, and the ensemble model was benchmarked against traditional daily life fraud detection methods through a comparative analysis. The findings revealed 99.89% accuracy of detection which is 5% higher than the previous studies [27].

Somkunwar et al. (2023) provided an innovative fraud detection system significantly improves the safety of financial networks. The study helps the banking sector by enabling the early detection of fraudulent transactions, which boosts confidence and security. With a fraud detection accuracy rate of 94.83%, this system effectively prevents financial crimes by utilizing machine learning and Benford Law [28].

Berkmans and Karthick's (2022) study indicated that the SMOTE-based sampling approach produces fruitful outcomes in the future. Using a DRF classifier allowed the SMOTE sampling strategy to get the maximum recall (0.81). This classifier achieved an accuracy score of 87.0%. Using all of the data that was gathered, the Stacked Ensemble algorithm came out on top with an average performance of 78.0%. The Stacked Ensemble has shown good performance in the majority of sampling operations as a fraud detection model [29].

The comparative analysis of background studies based on their Methodology, Data, Problem Addressed, Performance, and future work and limitations is provided in Table I.

TABLE I. Summery of Literature in Banking Systems using machine learning approaches

Author	Methodology	Dataset	Problem Addressed	Performance	Future Work / Limitation
Mishra, Biswal & Padhy (2025)	ML classifiers: LR, RF, SVM, KNN, GB, AdaBoost, DT	Banking cybersecurity dataset (unspecified)	Fraud detection in banking using ML classifiers	RF: Accuracy = 98.5%, Recall = 98.5%; KNN: F1-score = 98.89%	Focus on real-time deployment, adversarial attacks not considered
Souran & Shah (2025)	Hybrid model: XGBoost + Isolation Forest	Public financial transaction dataset	Reduce false positives and enhance fraud detection accuracy	Accuracy = 99.1%, AUC-ROC = 0.996	Requires robust anomaly update mechanisms for new patterns
R et al. (2025)	Stacking Ensemble Model (SEM) + Correlation Analysis	Credit Card Fraud Detection (CCFD) dataset (Kaggle)	Handling data drift, feature optimization, fraud detection	Accuracy = 99.8%, AUC = 93.5%	Future exploration of dynamic data drift adaptation

© JGREC 2026, All Rights Reserved

Keerthana et al. (2024)	Deep Learning: RNNs + CNNs	Real-time digital banking transaction data	Detecting real-time fraud in digital banking	Accuracy = 95.2%, Precision = 91.0%, Recall = 89.3%	Generalizability across varied banking environments
Narejo et al. (2024)	LSTM + SeqTrans hybrid ensemble	Daily banking fraud dataset (credit/debit/OBA)	Time-series-based fraud detection in banking	Accuracy = 99.89%	Model complexity and computational overhead
Somkunwar et al. (2023)	Benford's Law + ML anomaly detection	Financial transaction numerical data	Detect anomalies using digit-based patterns	Accuracy = 94.83%	Limitations in non-numerical fraud behaviour
Berkmans & Karthick (2022)	SMOTE + DRF + Stacked Ensemble	Credit Card Fraud Detection (CCFD) dataset	Dealing with unequal access to resources in fraud detection	DRF + SMOTE: Recall = 0.81, Accuracy = 87%; Stacked Ensemble Avg = 78%	Need for performance tuning and model optimization

III. Methodology

The research methodology for this study began with comprehensive data preprocessing on the IEEE CIS Fraud Detection dataset. The pre-processing included removing features with high percentages of missing values, imputing numerical data using median values, and normalizing features to a consistent scale using min-max normalization. A crucial step was feature selection, and SMOTE technique was applied to create a balanced representation of both normal and fraudulent transactions. Next, a 70/30 split of the data was made for testing and training. Ultimately, two models LightGBM and a Convolutional Neural Network (CNN) were trained and assessed using performance criteria. comprise the accuracy, F1-score, recall, and precision. The whole development steps are shown in Fig. 1.

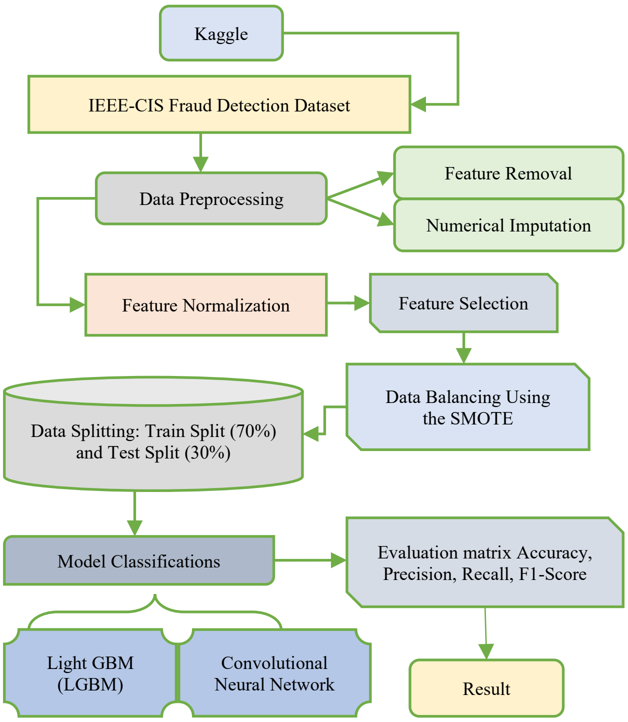
Kaggle
IEEE-CIS Fraud Detection Dataset
Data Preprocessing
Feature Removal
Numerical Imputation
Feature Normalization
Feature Selection
Data Balancing Using the SMOTE
Data Splitting: Train Split (70%) and Test Split (30%)
Model Classifications
Evaluation matrix Accuracy, Precision, Recall, F1-Score
Light GBM (LGBM)
Convolutional Neural Network
Result

Fig. 1. Flowchart of Financial Fraud Detection in Real-Time Banking Systems.

The following section provides a detailed explanation of each step in the flowchart:

A. Data Analysis

The IEEE CIS Fraud Detection dataset, which comprises financial transactions from Kaggle, was chosen for this investigation. It contains a large number of variables and data, including training and test subsets, however only uses the training data due to the lack of fraud incidence labels in the test data. It has 394 columns and 590,540 rows overall, and it includes details on the payer and merchant who completed the transaction. It has 144233 rows/data samples and 41 columns/features. Below are some data visualizations given on the IEEE CIS Fraud Detection dataset:

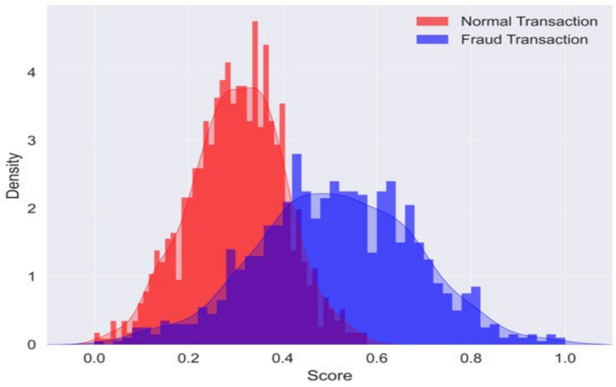

Fig. 2. KDE Plot for normal and Fraud score.

Fig. 2 distribution of transaction scores for "Normal Transaction" and "Fraud Transaction" is clearly separable. Normal transactions, shown in red, are heavily concentrated at lower scores, with a peak density around 0.3. In contrast, fraudulent transactions, depicted in blue, have a higher average score, with their distribution centered around 0.6. While there is a slight overlap in the middle, the distinct peaks and distributions suggest that a threshold could be set to effectively differentiate between normal and fraudulent transactions, allowing for accurate classification.

B. Data Preprocessing

In order to keep the input representations across various financial datasets and to clean the data for learning, pre-processing is a crucial step. There are various phases in the study's preparation pipeline: Feature Removal, Numerical Imputation, and feature normalization. These steps are discussed below:

- **Feature Removal:** Features that have more than 95% missing values are eliminated in order to avoid sparse representations that might deceive ensemble learners.
- **Numerical Imputation:** In numerical characteristics, class-specific distributions are maintained by using median imputation within fraud/legitimate groupings independently.

C. Feature Normalization

The technique of scaling numerical input variables to a common range or distribution is known as feature normalization, such as [0, 1] or the standard normal distribution, to improve model performance, convergence speed, and comparability across features. Min-max normalization is used to keep training from being dominated

© JGREC 2026, All Rights Reserved

by characteristics with bigger scales. Consider a raw feature value x_i∈R and its normalised version, x_i', As shown in Equation (1):

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

This change preserves proportional relationships by mapping all values to the interval [0, 1].

D. Feature selection

The process of Feature selection is selecting the most relevant subset of characteristics for an AI/ML model that works. This process reveals that while a few features are highly influential, the importance of others diminishes significantly, allowing for the removal of less-important features without a substantial drop in model accuracy.

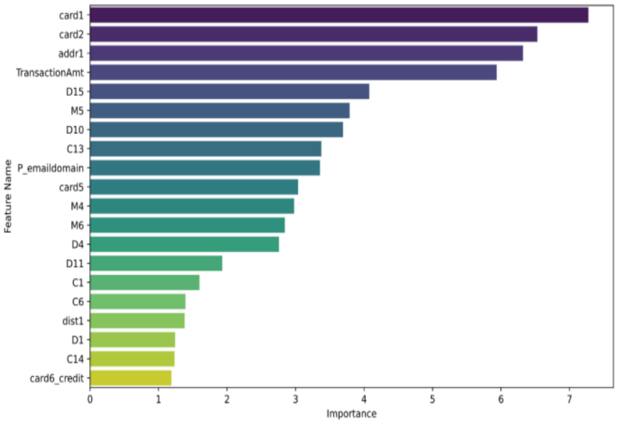

Fig. 3. The Top 20 features selected.

The top 20 most crucial characteristics chosen for the fraud detection model are shown in Fig. 3, arranged according to their significance ratings on a scale of 0 to 7. The feature "C4V3" emerges as the most important predictor, with a about 6.5 importance score, followed by "C4V1" and "C4V2" with scores around 6.0 and 5.5 respectively. The remaining features show a gradual decline in importance, with "TransactionAmt" (transaction amount) being notably prominent among the top features at approximately 5.0. A few features, like the C4V series, are highly influential. However, the importance drops significantly after the top five features. The remaining 15 features have modest but still meaningful contributions. Their importance scores range from 0.5 to 3.0.

E. Data Balancing with Synthetic Minority Over-sampling Technique (SMOTE)

In order to resolve the class imbalance present in datasets, improved the training procedure for traditional ML models by utilizing the SMOTE. Fig. 4 provides the bar graph for data balancing.

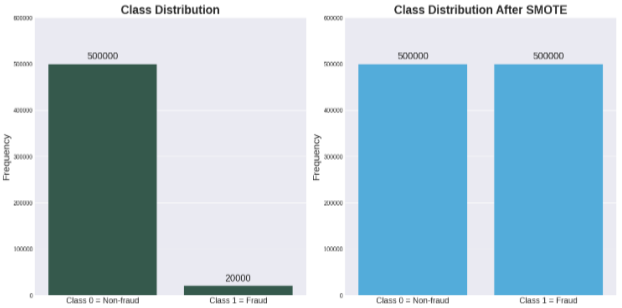

Fig. 4. Bar Graph for data distribution before and after balancing.

In Fig. 4, the class distribution of a dataset is seen both before and after SMOTE is used for data balancing. The left panel shows the original imbalanced dataset where Class 0 dominates with approximately 500,000 samples, while Class 1 represents a small minority with only about 20,000 samples, creating a severe class imbalance ratio of roughly 25:1. The right panel demonstrates the effectiveness of SMOTE in addressing this imbalance, showing that both classes now have equal representation with 500,000 samples each. For ML models to be trained successfully, this balanced distribution is essential.

F. Train-Test Split

Experimenting with both train/test splits showed that the 70/30 split appears to have a better representation of Training and testing datasets contain both positive and negative classes.

G. Propose Light GBM Model

The Light Gradient Boosting Machine, often referred to as LightGBM or LGBoost [30]. By employing a maximum depth restriction and a leaf-wise growth approach, this histogram-based technique speeds up training and lowers memory use. The leaves on the same layer are divided concurrently using the level-wise growth technique. Despite their varying information gains, leaves on the same layer undergo indiscriminate processing [31]. Information gain shows the anticipated decrease in entropy brought about by dividing the nodes according to characteristics as given in the Equations (2) and (3):

$$IG(B,V) = En(B) - \sum_{v\in(v)} \frac{|B_v|}{B} En(B_v) \quad (2)$$

$$En(B) = \sum_{d=1}^{D} -p_d \, log_2 \, p_d \quad (3)$$

Where p_d is the percentage of B that falls into category d, and D is the number of categories [32]. $En\ (B)$ is the collection's information entropy. The value of attribute V is denoted by B_v, and the subset of B for which attribute has value v is denoted by B_v.

H. Convolutional Neural Network:

CNNs are created to analyze structured data arrays in any format using deep learning [33]. The CNN architecture consists of three main levels The three types of layers are as follows (a) convolutional, (b) pooling, and (c) fully connected. There are input and output layers as well. The model's total performance is highly reliant on each layer [34], [35].

CNN differs from conventional neural networks due to its convolutional layer, a crucial feature layer composed of groups of convolutional filters. Convolution is transformed into a correlation operation by using a symmetric kernel. It extracts important features from the images and converts them into a small matrix in accordance with the kernel size [36]. The convolution operation equation is provided with no padding as shown in Equation. (4):

$$S_{i,j} = (I * K)i,j = \sum m \sum_n I_{i,j} \cdot K_{i-m,j-n} \quad (4)$$

Convolution creates a convolved picture by moving a tiny filter from top to bottom and left to right. This procedure is then repeated to generate several output feature maps.

I. Performance Matrix

A range of performance assessment criteria is used in this study to determine whether the research is successful [37]. These approaches rely on the confusion matrix that is generated throughout the identification job testing process. The following computations are used in these procedures:

- **True Positives (TP):** When a fraudulent transaction occurs, the system accurately identifies it as fraud.
- **True Negatives (TN):** The accurate categorization of a valid transaction is as non-fraudulent.
- **False Positives (FP):** An innocent purchase is mistakenly marked as fraudulent.
- **False Negatives (FN):** The false impression that a transaction is legitimate leads to fraud.

These values are computed for each class individually during the testing phase and form the basis for calculating the assessment measures. The matrix is formulated in (5) to (10):

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP+TN+FP+}FN)} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})} \tag{7}$$

$$\text{F}_1 - \text{Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{8}$$

$$TPR = \frac{TP}{TP+FN} \tag{9}$$

$$FPR = \frac{FP}{FP+TN} \tag{10}$$

Accuracy is a ratio of the rightly classified transactions. The precision is primarily focused on the correctness of optimistic forecasts. It is the rate of projected fraudulent transactions which actually become fraudulent. The model's recall indicates how well it can identify every instance of fraud. The harmonic mean of the accuracy and recall is the F1-Score. Lastly, the ROC curve is used to assess a binary classifier's performance. Additionally, the True Positive Rate and False Positive Rate are graphed at various threshold values.

IV. Results and Discussion

This study confirms that the DL and ML models are able to identify money laundering and fraud of banking systems in real-time. This necessitated the use of strong graphics processing unit (GPU) and memory (RAM) of 16 GB in combination with an NVIDIA GeForce RTX 3070 Ti laptop that would allow real-time analysis. The CNN model is more effective than the Light GBM model. The CNN model had a larger Accuracy (99.91% vs. 98.27%), which implies that the total error rate is much lower as can be seen in Table II. It also had a little higher Precision (98.65% vs. 98.4), and Recall (99.86% vs. 99.8%), which denoted that it had fewer false positives and was more effective in recognizing positive cases of fraud. The greater F1-Score (99.73% vs. 99.1%), which is a balanced measure of the model's efficacy in handling imbalanced fraud datasets and is generated from the harmonic mean of accuracy and recall, further validates the CNN model's superior performance.

TABLE II. Performance Evaluation of Financial Fraud Detection on IEEE CIS data

Metrics	Light GBM	CNN
Accuracy	98.27	99.91
Precision	98.4	98.65
Recall	99.8	99.86
F1-Score	99.1	99.73

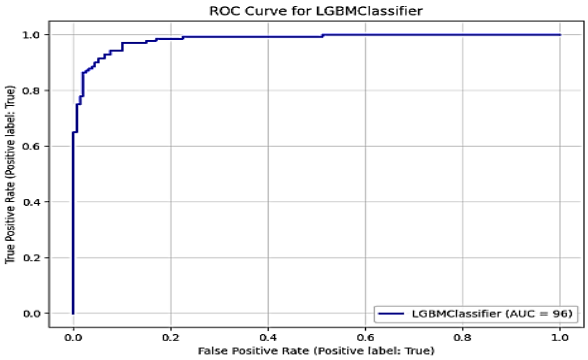

Fig. 5. The ROC Curve of the LGBM Classifier.

The trade-off between FPR and TP at various values is depicted by the LGBM Classifier ROC curve in Fig. 5. The AUC of 0.96 and the high climb of the curve towards the upper left corner show that the model is quite good at discriminating. It can use the ROC curve to evaluate balanced and unbalanced datasets because it performs well when it comes to class distinction.

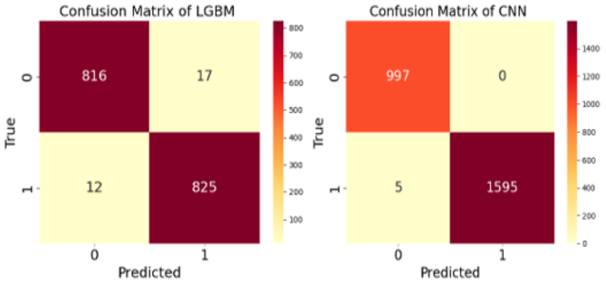

Fig. 6. Confusion Matrix of LGBM and CNN Model.

In Fig. 6 depicted confusion matrices for the LGBM and CNN models, the CNN model demonstrates superior classification performance, particularly in identifying true fraud cases. For the LGBM model, a small number of fraudulent transactions were misclassified as normal 17 legitimate transactions were mistakenly reported as fraudulent (False Positives), whereas 12 False Negatives were detected. In contrast, the CNN model was significantly more accurate, with only 5 FN and 0 FP. This implies that the CNN model is more reliable in identifying fraud as it avoids misclassifying normal transactions as fraudulent while simultaneously accurately identifying a greater number of real fraud occurrences.

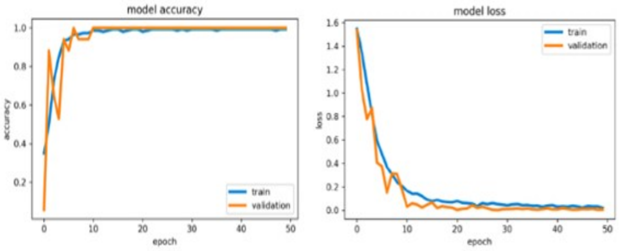

Fig. 7. Loss and Accuracy Graph for CNN Model.

Fig. 7 demonstrates strong performance with minimal overfitting during training. The training and validation accuracy curves on the "accuracy of model" graph both rise dramatically during the first ten epochs before stabilizing at a very high level, near 1.0. This suggests that the model is picking up new information efficiently and generalizing

© JGREC 2026, All Rights Reserved

successfully. The training and validation loss curves both rapidly decline before flattening out and remaining quite close to one another, which is further supported by the "model loss" graph. The training and validation curves for accuracy and loss exhibit almost similar behaviour, indicating that the model is well-suited for its task and is not memorizing the training data.

A. Comparative Analysis

A comparison of numerous well-known methods, based on their effectiveness in detecting fraudulent transactions, is presented in Table III below. CNN is the most successful model for financial fraud detection, with 99.91% accuracy, 98.65% precision, 99.86% recall, and 99.73% F1-score, according to the comparison research. LGBM ranks second with strong metrics (98.27% accuracy, 99.1% F1-score), while traditional algorithms like AdaBoost and Decision Tree show moderate performance at 92% accuracy. LSTM underperforms at 84.6% accuracy, and GNN demonstrates the weakest results at 78.01% accuracy. Notably, the Naive Bayes are inconsistent and has high precision (97.22%) and lower recall (84.87%), which may indicate overfitting. In general, CNN has better pattern recognition, which makes the tool the best in detecting financial fraud when compared with both standard ML and alternative DL models.

TABLE III. COMPARATIVE ANALYSIS OF FINANCIAL FRAUD DETECTION

Models	Accuracy	Precision	Recall	F1-score
LSTM[38]	84.6	84	85.7	84.8
DT[39]	92	92	93	90
GNN[40]	78.01	80.98	79.68	80.78
NB[41]	9122	97.22	84.87	90.63
Ada[42]	92	92	92	92
LGBM	98.27	98.4	99.8	99.1
CNN	99.91	98.65	99.86	99.73

The importance of the research is that it can offer a highly effective solution to identifying financial fraud is one of the most important issues facing the banking industry. The first advantage is the high accuracy to which the CNN model reaches an impressive standard of performance. This accuracy is not only keeping the financial institutions at a safe distance of incurring massive losses but also enhancing consumer confidence through the accurate identification of minimizing false positives and reducing false frauds. The systematic data preprocessing and application of SMOTE to the study also provide a useful and reproducible system to process an imbalanced dataset in other fields.

B. Justification and novelty

The proposed study offers an innovative method of identifying credit card fraud by combining sophisticated feature selection, SMOTE-based class balancing and ensemble learning. The reason why this work is justified is that the financial losses caused by fraud have been increasing, and advanced detection systems are needed that would be able to follow the changing dynamics. This is innovative in that it integrates the analysis of importance of features (key C4V series predictors) with synthetic minority oversampling to balance the learning process and ensemble to better accuracy and reduce false positives, resulting in a more reliable and effective way to identify fraud in real time.

V. CONCLUSION AND FUTURE SCOPE

Detection of financial fraud continues to be a big challenge to the contemporary financial institutions, especially on real time transaction that requires interpretability, speed and accuracy. LightGBM and CNN ML and DL methods were applied to IEEE-CIS Fraud Detection dataset in this study. Both effective pre-processing techniques including feature selection, data normalization, and data balance helped the models perform well during prediction. This research paper affirms that the proposed fraud detection CNN is effective because it is the best model given its high accuracy rates as it has the highest accuracy rate of 99.91. LGBM also was found to be powerful in capturing complex patterns, which is indicative of possible complementary use. The ability to conduct comparative analysis with other models proved that the suggested approaches are better, as they could be used in scalable and real-time banking applications. The primary weakness of the research is that it uses a static dataset, thus the model cannot perform in a dynamic setting in terms of fraud trends in a real-life, dynamic setting. The major issue is the generalizability of the model to the new, yet undetected frauds. To overcome this, future efforts should be on real-time deployment to test the performance under the conditions of the activity and on the development of novel methods to train the model on multiple, frequently updated data sets, such as federated learning or transfer learning, making it more adaptable to new fraud schemes and without violation of confidentiality of data.

REFERENCES

[1] N. Malali, "AI Ethics in Financial Services: A Global Perspective," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 2, 2025, doi: 10.5281/zenodo.14881349.

[2] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 10, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[3] V. Pal and S. K. Chintagunta, "Transformer-Based Graph Neural Networks for RealTime Fraud Detection in Blockchain Networks," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 3, pp. 1401–1411, Jul. 2023, doi: 10.48175/IJARSCT-11978Y.

[4] N. Malali, "Exploring Artificial Intelligence Models for Early Warning Systems with Systemic Risk Analysis in Finance," in *2025 International Conference on Advanced Computing Technologies (ICoACT)*, IEEE, Mar. 2025, pp. 1–6. doi: 10.1109/ICoACT63339.2025.11005357.

[5] Y. Macha and S. K. Pulichikkunnu, "An Explainable AI System for Fraud Identification in Insurance Claims via Machine-Learning Methods," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 3, pp. 1391–1400, Jul. 2023, doi: 10.48175/IJARSCT-11978X.

[6] K. B. Thakkar and H. P. Kapadia, "The Roadmap to Digital Transformation in Banking: Advancing Credit Card Fraud Detection with Hybrid Deep Learning Model," in *2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST)*, IEEE, Apr. 2025, pp. 1–6. doi: 10.1109/ICTEST64710.2025.11042822.

[7] D. Patel, "Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model," *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, Dec. 2023, doi: 10.14741/ijcet/v.13.6.10.

[8] V. Verma, "Deep Learning-Based Fraud Detection in Financial Transactions: A Case Study Using Real-Time Data Streams," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[9] S. B. Karri, S. Gawali, S. Rayankula, and P. Vankadara, "AI Chatbots in Banking: Transforming Customer Service and Operational Efficiency," in *Frontiers in Artificial Intelligence and ApplicationsEbookVolume 414: Advancements in Smart Innovations, Intelligent Systems, and Technologies*, 2025, pp. 61–81. doi: 10.3233/FAIA251498.

[10] H. Kali, "Optimizing Credit Card Fraud Transactions Identification and Classification in Banking Industry Using Machine Learning Algorithms," *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 11, pp. 85–96, 2024.

[11] B. Chaudhari and S. C. G. Verma, "Synergizing Generative AI and

Machine Learning for Financial Credit Risk Forecasting and Code Auditing," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 2, 2025.

[12] V. Verma, "Security Compliance and Risk Management in AI-Driven Financial Transactions," *Int. J. Eng. Sci. Math.*, vol. 12, no. 7, pp. 1–15, 2023.

[13] B. R. Ande, "Federated Learning and Explainable AI for Decentralized Fraud Detection in Financial Systems," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 35s, pp. 48–56, 2025.

[14] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, pp. 16–22, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[15] R. Jain, S. K. Das, and Y. Makin, "Behavioral Risk Tolerance in U.S. Retirement Planning vs. Property Insurance: A Comparative Analysis," *Int. J. Appl. Math.*, vol. 38, no. 4s, pp. 41–70, Sep. 2025, doi: 10.12732/ijam.v38i4s.215.

[16] R. Dattangire, R. Vaidya, D. Biradar, and A. Joon, "Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality," in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ACET61898.2024.10730334.

[17] P. Chandrashekar, "Data-Driven Loan Default Prediction: Enhancing Business Process Workflows with Machine Learning," *Int. J. Emerg. Res. Eng. Technol.*, vol. 6, no. 4, 2025, doi: 10.63282/3050-922X.IJERET-V6I4P103.

[18] S. K. Tiwari, "The Future of Digital Retirement Solutions: A Study of Sustainability and Scalability in Financial Planning Tools," *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 5, pp. 229–245, Dec. 2024, doi: 10.32996/jcsts.2024.6.5.19.

[19] S. J. Wawge, "A Survey on the Identification of Credit Card Fraud Using Machine Learning with Precision, Performance, and Challenges," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3345–3352, May 2025, doi: 10.38124/ijisrt/25apr1813.

[20] H. Yaseen and A. Al-Amarneh, "Adoption of Artificial Intelligence-Driven Fraud Detection in Banking: The Role of Trust, Transparency, and Fairness Perception in Financial Institutions in the United Arab Emirates and Qatar," *J. Risk Financ. Manag.*, vol. 18, no. 4, 2025, doi: 10.3390/jrfm18040217.

[21] P. R. Marapatla, "AI-driven donor management: revolutionizing nonprofit fundraising through predictive analytics," *Int. J. Res. Comput. Appl. Inf. Technol.*, vol. 8, no. 1, 2025.

[22] A. R. Bilipelli, "Forecasting the Evolution of Cyber Attacks in FinTech Using Transformer-Based Time Series Models," *Int. J. Res. Anal. Rev.*, vol. 10, no. 3, pp. 383–389, 2023.

[23] J. Mishra, B. B. Biswal, and N. Padhy, "Machine Learning for Fraud Detection in Banking Cybersecurity Performance Evaluation of Classifiers and Their Real-Time Scalability," in *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, IEEE, Feb. 2025, pp. 431–436. doi: 10.1109/ESIC64052.2025.10962752.

[24] M. Souran and R. S. Shah, "Fraud Detection in Mobile Payments Using a Hybrid Machine Learning Model," in *2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*, 2025, pp. 1–6. doi: 10.1109/AMATHE65477.2025.11081192.

[25] R. V. S. S. B. R, H. MuhamedAle, C. Supriya, S. Ajay, and S. Kaliappan, "Automated Financial Fraud Detection in Banking Systems based on Stacking Ensemble Model," in *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, IEEE, Apr. 2025, pp. 1–5. doi: 10.1109/ICDCECE65353.2025.11035889.

[26] V. Keerthana, S. M, L. A, and P. K, "Adaptive Fraud Detection in Digital Banking Using Deep Learning: A Hybrid CNN-RNN Approach," in *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, IEEE, Dec. 2024, pp. 1–7. doi: 10.1109/ICSES63760.2024.10910880.

[27] A. Narejo, J. P. Li, A. N. Sanjrani, A. A. Sanjrani, and A. Iqtidar, "Dynamic Temporal LSTM- Seqtrans for Long Sequence: An Approach for Credit Card and Banking Accounts Fraud Detection in Banking System," in *2024 21st International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2024, pp. 1–10. doi: 10.1109/ICCWAMTIP64812.2024.10873619.

[28] R. K. Somkunwar, M. P. Pimpalkar, K. M. Katakdound, A. S. Bhide, S. P. Chinchalkar, and Y. M. Patil, "A Fraud Detection System in Financial Networks Using AntiBenford Subgraphs and Machine Learning Algorithms," in *IEEE 1st International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics, AIKIIE 2023*, 2023. doi: 10.1109/AIKIIE60097.2023.10390325.

[29] T. J. Berkmans and S. Karthick, "Credit Card Fraud Detection with Data Sampling," in *2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, 2022, pp. 1–6. doi: 10.1109/ICPECTS56089.2022.10046729.

[30] S. R. Kurakula, "Designing Enterprise Systems for the Future of Financial Services: The Intersection of AI, Cloud-Native Microservices, and Intelligent Data Processing," *Eur. Cent. Res. Train. Dev. UK*, vol. 13, no. 20, pp. 91–103, 2025.

[31] R. Q. Majumder, "Machine Learning for Predictive Analytics: Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3557–3564, May 2025, doi: 10.38124/ijisrt/25apr1899.

[32] R. Szczepanek, "Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost," *Hydrology*, vol. 9, no. 12, p. 226, Dec. 2022, doi: 10.3390/hydrology9120226.

[33] T. V. Shah, "Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing," *Int. J. Sci. Res. Arch.*, vol. 15, no. 3, pp. 618–627, Jun. 2025, doi: 10.30574/ijsra.2025.15.3.1767.

[34] S. Mathur and S. Gupta, "Classification and Detection of Automated Facial Mask to COVID-19 based on Deep CNN Model," in *2023 IEEE 7th Conference on Information and Communication Technology (CICT)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/CICT59886.2023.10455699.

[35] S. Nokhwal, P. Chilakalapudi, P. Donekal, S. Nokhwal, S. Pahune, and A. Chaudhary, "Accelerating Neural Network Training: A Brief Review," in *2024 8th International Conference on Intelligent Systems Metaheuristics & Swarm Intelligence (ISMSI)*, New York, NY, USA: ACM, Apr. 2024, pp. 31–35. doi: 10.1145/3665065.3665071.

[36] M. A. K. Raiaan *et al.*, "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," *Decis. Anal. J.*, vol. 11, p. 100470, Jun. 2024, doi: 10.1016/j.dajour.2024.100470.

[37] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

[38] A. H. Almuteer, A. A. Aloufi, W. O. Alrashidi, J. F. Alshobaili, and D. M. Ibrahim, "Detecting Credit Card Fraud using Machine Learning," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 24, pp. 108–122, Dec. 2021, doi: 10.3991/ijim.v15i24.27355.

[39] J. K. Afriyie *et al.*, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decis. Anal. J.*, vol. 6, p. 100163, Mar. 2023, doi: 10.1016/j.dajour.2023.100163.

[40] F. K. Alarfaj and S. Shahzadi, "Enhancing Fraud Detection in Banking With Deep Learning: Graph Neural Networks and Autoencoders for Real-Time Credit Card Fraud Prevention," *IEEE Access*, vol. 13, pp. 20633–20646, 2025, doi: 10.1109/ACCESS.2024.3466288.

[41] B. Borketey, "Real-Time Fraud Detection Using Machine Learning," *J. Data Anal. Inf. Process.*, vol. 12, no. 02, pp. 189–209, 2024, doi: 10.4236/jdaip.2024.122011.

[42] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers," *Risks*, vol. 12, no. 11, p. 174, Nov. 2024, doi: 10.3390/risks12110174.

© JGREC 2026, All Rights Reserved