Journal of Global Trends in Artificial Intelligence (JGTAI)

Volume 1, No. 1, June 2025
Available Online at: www.jgtai.com

An Advanced Machine Learning Method Used for Identification of Banking Fraud in Financial Transactions

Sandeep Gupta
SATI, Vidisha
Sandeepguptabashu@gmail.com

***Abstract*—In financial security, identifying credit card fraud is still a major problem because of the wildly disproportionate size of transaction databases and the constantly changing strategies used by criminals. To identify credit card fraud, this study uses a Convolutional Cascaded Neural Network (CCNN) model on a highly imbalanced data set (284,807 anonymized transactions) to get reliable results. The data is subjected to a thorough preparation procedure that addresses missing values, uses PCA to extract features, encodes labels, and uses SMOTE to solve the imbalance of minority classes. The CCNN model is trained to determine if the transactions are genuine or fraudulent once the data is divided into training (70%) and testing (30%) sets. Experiments show that the CCNN significantly outperforms well-known models like Naive Bayes (NB) and K-Nearest Neighbors (KNN), thanks to its remarkable 99.93 percent accuracy, 99.85 percent precision, 99.99 percent recall, and 99.93 percent F1-score. The model demonstrated the potential to be applied as a scalable and reliable solution to the real-world credit card fraud prevention situation, and these results supported the model in offering reliable results in detecting fraudulent behavior with very little false positives and false negatives.**

Keywords—Credit Card Fraud Detection, Banking, Continuous Convolutional Neural Network (CCNN), Imbalanced Dataset, Principal Component Analysis (PCA), Deep Learning, Financial Transactions, Classification, Evaluation Metrics.

I. INTRODUCTION

Financial institution fraud, also known as bank fraud, encompasses a wide range of crimes committed against banks and their clients. Deceptive practices used to illegally acquire funds or assets from a bank are known as bank fraud [1][2][3][4]. Financial institutions can analyze vast volumes of transactional data in real time, protect assets, and maintain client confidence by responding swiftly to fraudulent events. Over the past three decades, financial institution frauds have been rising exponentially, resulting in significant annual financial losses for banks [5][6][7]. To mitigate these increasing losses, banks are developing more sophisticated anti-fraud programs and detection systems.

Fraud detection in financial transactions relies heavily on advanced analytics, ML algorithms, and AI techniques to identify anomalous patterns or behaviors indicative of fraudulent activities [8][9][10]. Financial institutions can analyze vast volumes of transactional data in real time, protect assets, and maintain client confidence by responding swiftly to fraudulent events [11]. However, a number of issues that these Credit Card Fraud Detection Systems (CCFDS) encounter, including idea drift, verification delay, and dataset class imbalance, make it more difficult to detect fraud effectively and promptly.

Artificial intelligence (AI), which encompasses robots' capacity to make decisions, solve problems, and recognize patterns, has changed several sectors, including corporate finance, which uses AI for risk management and fraud detection [12]. ML, a branch of AI, develops methods to teach computers new skills and improve their performance as they adjust to changes without the assistance of humans or code [13][14][15][16]. Contrary to the static rule-based systems, the ML algorithms are constantly refining their detection rates and effectiveness by learning new data input and identifying previously unrecognized patterns.

Fraudulent transactions have become more and more complex and their volumes have increased to the extent that human or rule-driven systems are not effective in detecting them. Consequently, ML technologies have become one of the possible solutions in the problem of detecting financial fraud [17]. Some of the benefits of ML-based solutions are that they can accommodate large volumes of data, and these systems are able to evolve to new fraud patterns as they come along and can correlate false positives through constant learning [18][19][20][21]. Not only does this assist the financial institutions to prevent financial losses, it also increases the experience of the customer by eliminating unwarranted transaction rejections [22][23][24][25]. Therefore, the adoption of ML methodologies in the fraud detection platforms is vital in achieving robust, scalable, and intelligent banking security systems that can be used to predict and fight the emergent fraud schemes.

A. Significance and Contribution

In this study, the significance of the use of advanced ML to enhance financial security through better detection of credit card theft is demonstrated. On a real data volume that is highly unbalanced, the study addresses prevalent problems of high-dimensional data and skewed class distribution using a CCNN. Synergy between a good preprocessing method and a deep learning model makes the proposed system highly precise and with minimal false detections, which makes the system practical and scalable in terms of real-time fraud prevention. The key contributions are the following:

- The research uses a real life credit card fraud detection dataset which has 284,807 records of transaction history.

- Performed extensive preprocessing involving the treatment of missing values, encoding of categorical variables and dimensionality reduction using PCA.
- A new form of a Convolutional Neural Network, termed Continuous Convolutional Neural Network (CCNN) is trained to recognize deep, non-linear transaction pattern, which provides better ability compared to the older models in recognizing and categorizing fraudulent and non-fraudulent activity.
- Industry-standard measures including accuracy, precision, recall, and F1-score are used in rigorous testing to demonstrate the model's efficacy and resilience in the real-time fraud detection use case.

B. *Justification and Novelty*

This thesis is based on the observation that predicting credit card theft by combining sophisticated preprocessing methods with a potent DL model is a novel and successful strategy. This is mainly because the normal models are not good enough to handle complex, non-linear trends in transaction data or highly skewed fraud datasets. To address these challenges, PCA is utilized for anonymized feature extraction, ensuring data privacy, while SMOTE is employed to balance the dataset, enabling more effective learning. The adoption of a CCNN introduces a novel architecture capable of capturing deep and complex relationships in the data, significantly enhancing detection accuracy. The proposed CCNN demonstrates its distinctiveness and use for identifying financial fraud in the actual world by outperforming conventional machine learning models on critical metrics including accuracy, precision, recall, and F1-score.

C. *Structure of the Paper*

Here is how the study is organized: In Section I, it covers the history, goals, and rationale of credit card fraud detection by ML. Section II provides a literature review that emphasizes new developments and areas where research is lacking. Section III describes the proposed methodology, including dataset description, data preprocessing techniques, model architecture, and evaluation metrics. Section IV discusses the experimental setup, results, and comparative analysis of the CCNN model with traditional classifiers. Section V concludes the study and outlines future directions.

II. Literature Review

This section discusses some review articles on CCFD using ML and DL approaches. Various studies have explored different models, preprocessing techniques, and performance metrics to improve fraud detection accuracy. Each paper's methodology, dataset, main conclusions, limitations, and recommendations for further research are highlighted in Table I.

Mishra, Biswal and Padhy (2025) used several ML classifiers to detect fraudulent behaviour in the banking system. They used classifiers such as LR, RF, SVM, KNN, GB, AdaBoost, and DT. They have estimated these classifiers 'performance metrics like: - accuracy, precision, recall, and F1-score. Result: From the experimental observation, they found that RF gives the highest accuracy of 0.985, Recall at 0.985000 and the classifier KNN has the highest score at 0.988937 [26].

Nair et al. (2025) examine how well data balancing strategies and ensemble learning approaches can identify CCF in unbalanced datasets. In addition to evaluating data balancing strategies like SMOTE, ADASYN, and Random Oversampling, they examine four ML algorithms: KNN, DT, SVM, and LSTM. Significant gains were made by using balancing techniques: The greatest results are obtained with ROS on LSTM, with Recall 89.6%, F1-Score 91.5% and Precision 98.4%, With Precision 97.6%, Recall 91.5%, and F1-Score 93.9%, Random Oversampling ranks second on SVM, whereas SWOT on SVM is third with Precision 91.7%, Recall 86.3%, and F1-Score 89.9 % [27].

Dharma and Latha (2025) CCFD can be solved by using ML techniques, when required data is gathered and available. The hybrid ML strategy for credit card fraud detection is shown in this research. The data used in this study came from 284,807 cardholder transactions made in Europe in September 2013. The hybrid ML strategy for credit card fraud detection is shown in this research. The information utilized in this investigation was gathered from 284,807 European cardholder transactions in September 2013. For developing hybrid model ML techniques like SVM and LR are used. The results of Hybrid ML methods are depended on Accuracy, Precision, Recall, and F1score. Results states that, described model achieves high performance parameters accuracy, Precision, Recall and F1score of 97%,96%,97%,970/0 respectively than other models [28].

Singh et al. (2024) compare the performance of two well-known ML methods, such as XGBoost and ANNs in identifying credit card fraud. They assess these algorithms' accuracy, precision, recall, and F1 score using a publicly available dataset of credit card transactions. The study also examines how well ANNs and XGBoost scale computationally in order to further evaluate their potential for application in real-time fraud detection systems. ANNs achieve the highest accuracy at 96.9%, surpassing all five methods evaluated, while XGBoost, with an accuracy of 92.7%, outperforms all other classifiers. These results provide financial organizations looking to create or improve fraud detection systems with useful information and highlight the benefits and drawbacks of each approach [29].

Jain et al. (2024) uses ML algorithms to identify financial transactions that are fraudulent by utilizing the Pay Sim dataset. In fraud detection tasks, the inherent class imbalance is addressed via label encoding, min-max scaling, and SMOTE-based balancing. SVM, NB, RF, and LR are examples of classification models that are evaluated using F1-score, recall, accuracy, and precision. The outcomes indicate that LR excels in accuracy with 98.99%, whereas RF excels in recall, and NB excels in precision. In order to tackle developing financial crimes, the study highlights the necessity of robust fraud detection systems as well as the trade-offs associated with optimizing evaluation techniques [30].

Silvia et al. (2024) examine how user behaviour patterns enhance fraud detection by integrating behavioral analytics with ML models to improve accuracy and reduce FP. A bibliometric analysis using VOS viewer was conducted on 200 Google Scholar papers (2020–2024) to identify key trends. Terms like “user behaviour” and “ML algorithms” emerged as central. ML models, including NN and DL, were evaluated, with CNN achieving up to 95% accuracy. Problems like overfitting, data privacy, and computational complexity still exist. Further investigation on hybrid models that combine various data sources to enhance fraud detection systems is needed [31].

Islam et al. (2023) in order to identify instances of bank fraud, the authors used a synthetic dataset of 1,00,000 rows and 32 columns, as well as four supervised machine learning models: KNN, RF, DT, and LR algorithms. This study was easy to read thanks to the manner it was pre-processed, decoded, feature-engineered, validated, evaluated, and explained. The area under the curve (AUC) for gradient boosting models stays quite high at 98%. Financial institutions such as banks and insurance providers may use this study in the real world [32] .

Although prior studies have employed machine learning models like RF, SVM, KNN, and XGBoost limitations in addressing class inequality and identifying intricate fraud patterns continue to exist in the detection of credit card fraud. Traditional models often underperform on highly imbalanced datasets, leading to high false positives or negatives. Techniques like SMOTE and ADASYN offer partial relief but are model-dependent. Moreover, deep learning models, particularly CNN-based architectures, remain underexplored in this domain. This study addresses these gaps by proposing a CCNN with PCA-based feature extraction and SMOTE balancing. The suggested paradigm improves real-time fraud detection in financial systems in terms of accuracy, resilience, and scalability.

TABLE I. SUMMARY OF BACKGROUND STUDY FOR CREDIT CARD FRAUD DETECTION

Author(s)	Methods Used	Dataset Used	Key Findings	Limitations & Future Work
Mishra, Biswal & Padhy (2025)	Random Forest, SVM, KNN, AdaBoost, Gradient Boosting, Logistic Regression, and Decision Tree	Not specified (banking system data)	Random Forest (RF) achieved highest accuracy (0.985) and recall. KNN had the highest score (0.988937). AdaBoost and Gradient Boosting showed strong precision and AUC-ROC.	Need for more real-world datasets and testing model robustness in dynamic banking environments.
Nair et al. (2025)	KNN, Decision Tree, SVM, LSTM + SMOTE, ADASYN, Random Oversampling	Unbalanced credit card fraud dataset	LSTM with ROS: Precision (98.4%), Recall (89.6%), F1-Score (91.5%). SVM with ROS: F1-Score (93.9%). Oversampling improves performance significantly.	Requires computational resources; future work may involve hybrid models or cost-sensitive learning for imbalance handling.
Dharma & Latha (2025)	Hybrid model (SVM + Logistic Regression)	European cardholder dataset (284,807 transactions, Sept 2013)	High performance was attained: F1-Score (97%), Accuracy (97%), Precision (96%), and Recall (97%).	Further optimization and testing needed on recent data; integration with real-time fraud detection systems suggested.
Singh et al. (2024)	XGBoost, Artificial Neural Networks (ANNs)	Public credit card transaction dataset	ANN achieved the highest accuracy (96.9%), followed by XGBoost (92.7%). ANNs demonstrated better scalability for real-time fraud detection.	ANNs are computationally intensive; future work should assess energy-efficient models and real-time scalability.
Jain et al. (2024)	SVM, Naïve Bayes (NB), Random Forest, Logistic Regression + SMOTE, Label Encoding, Scaling	PaySim dataset	Logistic Regression had highest accuracy (98.99%), Random Forest best in recall, NB in precision. Showcased importance of preprocessing and model-specific strength.	Dataset-specific results; future work to explore ensemble/hybrid models across diverse financial datasets.
Silvia et al. (2024)	CNN, Deep Learning models + User Behavior Analytics + Bibliometric Analysis	Behavioral data + bibliometric data (Google Scholar papers 2020–2024)	CNN models reached up to 95% accuracy. User behavior analytics improved fraud detection precision.	High computational cost, overfitting risks, and data privacy issues remain; future work should explore multi-source data integration.
Islam et al. (2023)	Decision trees, KNN, Random Forest, Gradient Boosting, and Logistic Regression	Synthetic dataset (100,000 rows, 32 columns)	Achieved 98% AUC with Gradient Boosting. Emphasized rigorous preprocessing and evaluation methods.	Relies on synthetic data; future work should validate models on real-life financial datasets and extend to insurance applications.

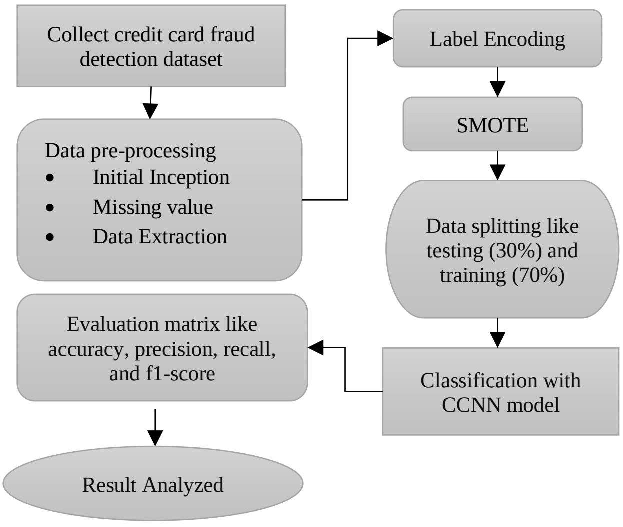
Collect credit card fraud detection dataset
Data pre-processing
Initial Inception
Missing value
Data Extraction
Label Encoding
SMOTE
Data splitting like testing (30%) and training (70%)
Classification with CCNN model
Evaluation matrix like accuracy, precision, recall, and f1-score
Result Analyzed

Fig. 1. Flowchart of the Banking Transaction Detection

III. METHODOLOGY

The suggested approach for identifying financial fraud in banking transactions, as shown in Figure 1, begins with the acquisition of a relevant dataset containing historical transaction records. To improve the calibre and applicability of inputs for the model, the data is subjected to thorough preprocessing, which includes initial inspection, treatment of missing values, and feature extraction. Categorical features are transformed into numerical format through label encoding to ensure compatibility with machine learning algorithms. In fraud datasets, the class imbalance problem is addressed by the SMOTE, which creates instances of the minority class—the fraudulent ones. The balanced dataset is separated into training (70%) and testing (30%) subsets for the purpose of building and evaluating the model. Transactions are classified as either legitimate or fraudulent using a Convolutional Cascaded Neural Network (CCNN). In real-world financial fraud detection scenarios, assessment measures including recall, accuracy, precision, and F1-score may be used to assess the model's efficacy and performance.

The following steps of proposed methodology are briefly discussing in below:

A. *Data Collection*

There were 284,807 transactions from European cardholders in the credit card fraud detection dataset over two days in September 2013, with a very low rate of 0.17% tagged as fraudulent. Each entry includes 30 numerical features 28 anonymized via PCA, along with 'Time', 'Amount', and a binary 'Class' label. The dataset is widely used to develop and evaluate fraud detection models in highly imbalanced, real-world scenarios.

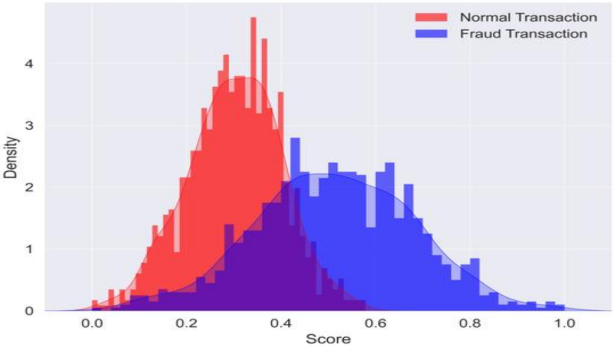

Fig. 2. Kernel Density Curve of Fraud Score

Figure 2 illustrate that it is a density plot comparing the distribution of scores for normal and fraud transactions. Normal transactions are shown by the red curve, whereas fraudulent transactions are shown by the blue curve. The score is displayed on the density is shown on the y-axis, and the x-axis. Normal transactions cluster around lower scores, while fraud transactions are more prevalent at higher scores. This visualization effectively highlights the separability between the two transaction types based on scoring.

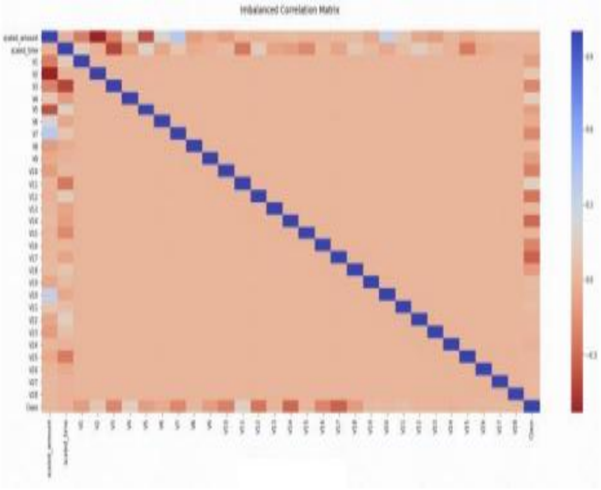

Fig. 3. Imbalanced Correlation Matrix

Figure 3 shows a correlation matrix heatmap titled "Imbalanced Correlation Matrix", representing the relationships among features in a CCFD dataset. The correlation coefficient between two variables is shown in each cell, with the intensity of the color representing the strength and direction of the association. A strong blue connection, a weak red correlation, or no correlation at all is represented by a neutral tone. The strong diagonal line reflects perfect self-correlation of each feature with itself.

B. *Data Preprocessing*

The identification of credit card fraud during pre-processing draws attention to the inherent noise and inconsistencies in the information. Initial data examination, dropna() handling of missing values, and deleting unnecessary columns are important tasks. Dimensionality reduction was performed via clustering to reduce categorical feature complexity. These steps follow standard practices from existing literature to enhance model performance and generalizability:

- **Initial Inspection:** The dataset was initially examined using the. head (), .info (), and. describe () functions to gain insights into its structure, data types, and summary statistics.
- **Missing values:** The initial stage of data cleansing involves dealing with missing values. Using the isnull(), missing data was found. sum () function. Records containing null values were removed using the dropna () method to maintain data integrity.
- **Feature Extraction:** The dataset is anonymized, and Principal Component Analysis (PCA) is used to extract characteristics. PCA is used to determine that 28 of the 30 features are major components (V1–V28). This dimensionality reduction technique helps retain the most significant variance in the data while ensuring confidentiality and reducing noise, which improves model performance and computational efficiency.

C. *Data Blancing with Synthetic Minority Over-sampling Technique (SMOTE)*

In classification jobs, class imbalance is lessened by using a data preparation method known as SMOTE. As illustrated in Figure 4, illustrates the effect of applying the SMOTE on an imbalanced credit card fraud dataset. The left chart (a) represents the original dataset, where normal transactions dominate at 99.83%, while fraudulent transactions make up only 0.17%. In contrast, the right chart (b) shows the dataset after SMOTE has been applied, resulting in a balanced distribution with 50% normal and 50% fraudulent transactions. This change is essential for enhancing machine learning models' performance since it guarantees that the minority class—fraud—is fairly represented throughout training.

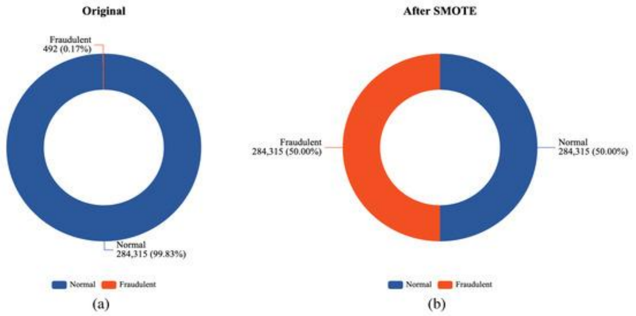

Fig. 4. Balancing with SMOTE

D. *Label Encoding*

Categorical features were encoded into numerical format utilizing the Label Encoder program from the Scikit-learn toolkit. For efficient model training and assessment, this transformation was required to turn non-numeric input into a machine-readable format.

E. *Data Splitting*

There are two subsets in the training and testing subsets of the credit card fraud detection dataset. The dataset is used for training 70% of the time and for testing 30% of the time.

F. *Classification with Continuous Convolutional Neural Network (CCNN) Model*

An updated neural network model that evolved from the PCNN is the CCNN. The objective of its design was to make it behave more like genuine neurones in response to dynamic stimuli. Essentially a multilayer perceptron, CNN is a specific form of feed-forward neural networks. Information pertaining to landslip vulnerability assessment is often presented in a one-dimensional format. Consequently, this study builds a 1D CNN model. Convolutional, pooling, and fully connected layers comprise the majority of the components in this model. The convolution layer convolves a series of factor vectors, and the maximum pooling layer pools them before the fully connected layer transfers the low-dimensional feature space using the high-dimensional feature information. Their last step is to use a nonlinear activation function to get the findings and probabilities for classifying landslides and non-landslides.

The convolutional layer performs feature extraction and feature mapping on the input data using a convolutional kernel. Equation (1) displays the formula for convolution:

$$X_j = f\left(\sum_i^N w_j \cdot x_i + b_j\right), j = 1,2,\cdots,k, \tag{1}$$

where w_j and b_j are the weight and bias coefficients, X_i is the local input data, X_j is the output result, and f is the nonlinear activation function. It is common practice to decrease the output dimensionality after the convolution procedure.

Overfitting and dimension explosion are problems that may arise when the convolutional module is directly connected to the classification layer. This issue is often addressed by connecting the pooling layer to the feature sampling convolutional layer. The CNN is able to adapt to subtle local morphological changes without undergoing distortion as a result of this reduction in parameter count and feature dimensionality. The research here makes use of the greatest pooling layer. Equation (2) shows the formula for calculating the maximum pooling layer:

$$p_i = \max_{i \in N}(a_i) \tag{2}$$

where pi is the outcome of the pooling procedure. The pooling window-corresponding input data is denoted by a_i, whereas i is the pooling location.

G. *Evaluation Metrics*

The last stage in assessing if the outcomes match the issue at hand is to look at how well the metrics included in ML models function. The metrics, which enable comparing the performance of models, show the capacity to accomplish a certain job, such as classification, regression, or clustering quality [33]. A true positive (TP) is a precise prediction of a positive value (fraud). A false negative (FN) would display a predicted negative with a positive real value, suggesting fraud; a false positive (FP) would display a projected positive with a negative actual value; and a true negative (TN) prediction would correctly display a negative result without fraud. The letters FP and FN stand for the misclassification cost, which is sometimes referred to as the prediction error of the classification model. The effectiveness of supervised machine learning techniques is evaluated using the following metrics:

1) *Accuracy*

The accuracy may be defined as the proportion of FP to FN in a set of fraudulent predictions. The precision metric is computed using Equation (3).

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+Fp+TN+FN}} \tag{3}$$

2) *Precision*

"What number of selected data items are relevant" is all it just highlights. Basically, it want to know how many of the observations that an algorithm has forecasted as positive really are positive. A system's accuracy may be defined as the ratio of its TP to its total positives and FP, as shown in Equation (4).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{4}$$

3) *Recall*

The accuracy metric known as recall TP contrasts the percentage of FP with the total number of erroneous negatives. To compute the sensitivity measure, use Equation (5).

$$Recall = \frac{\mathrm{TP}}{TP+FN} \tag{5}$$

4) *F1-score*

This measure, which may be referred to by other names like f-score or f-measure, accounts for both recall and precision when determining an algorithm's performance. This mathematical expression, given by Equation (6), represents the precision and recall harmonic mean:

$$F1 - Score = \frac{2(Precision*Recall)}{\mathrm{Precision+Recall}} \tag{6}$$

The machine and DL models are developed using these matrices.

IV. Result Analysis And Discussion

This section gives the experimental analysis of an algorithm that identifies fraudulent credit card transactions based on the Convolutional Neural Network (CCNN) model. The model was trained and tested using the publicly accessible Credit Card Fraud Detection dataset. The same system with a 3.3 GHz Intel dual-core i6 processor, 1 TB of RAM and the Windows 10 operating system was used to perform all the experiments. Table II indicates the excellent work of a CCNN model as it reached 99.93 percent of accuracy, which means almost perfect classification. The model has a 99.85 precision, which is minimal FP, and 99.99 recall that makes it not to miss any TP. Its balanced precision and recall are proved again with the 99.93% F1-score and therefore this model is very reliable when it comes to a high-performance classification task like a medical image or defect detection.

TABLE II. Performance Metrics of the CCNN Model

Performance Measures	CCNN
Accuracy	99.93
Precision	99.85
Recall	99.99
F1-score	99.93

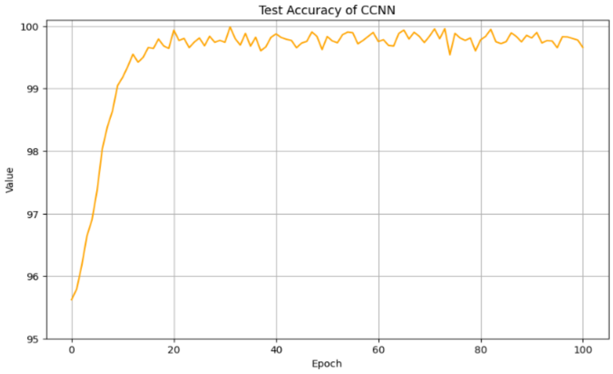

Fig. 5. Accuracy Graph of CCNN Model

The evolution of the CCNN model's test accuracy over the course of 100 training epochs is seen in Figure 5. The accuracy values are on the y-axis, whilst the x-axis shows the number of epochs. The model depicts fast growth of accuracy at the very beginning of the training process with fast decrease to the constant high level. This tendency indicates high learning and generalization ability of the model, which preconditions its applicability to such precision-oriented tasks as fraud detection.

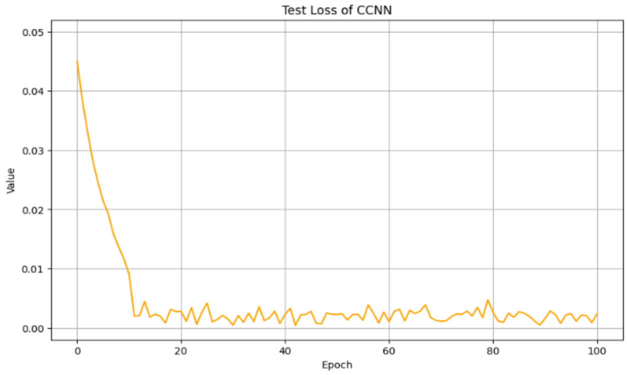

Fig. 6. Loss Graph of CCNN model

The test loss outcomes of the CCNN model in relation to training epochs are shown in Figure 6. The graph shows a quick decline of test loss in the first epoch of about 0.045 to the stability and low levels of test loss below 0.005 within the first 20 epochs. This is the lowest value in test loss and after this point, there are slight variations of the test loss around this low value, which is a sign of good model convergence and generalization over the rest of the training iterations up to epoch 100.

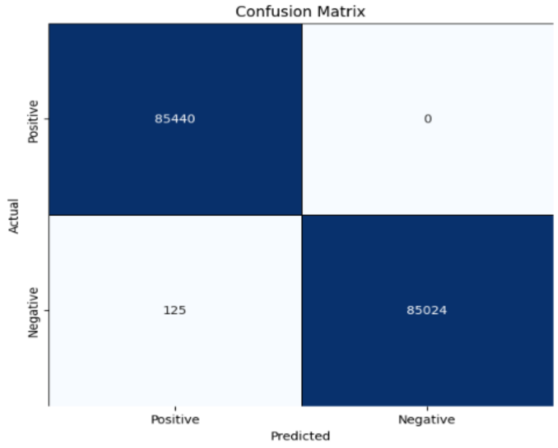

Fig. 7. Confusion matrix of CCNN model

The confusion matrix in figure 7 illustrates how well the CCNN model performed on the test dataset. With 125 FP and no FN, it shows 85,440 true positives and 85,024 genuine negatives. These results confirm that the ability of The model that accurately categorizes illegal transactions and prevent false positives of legitimate transactions was great, demonstrating its effectiveness and dependability in identifying credit card frauds in the real world.

A. Comparative Analysis and Discussion

The comparative examination of classification performance between the CCNN, NB, and KNN models is presented in this section. Accuracy, precision, recall, and F1-score all crucial performance metrics are used for the comparison. On the CCFD dataset, Table III presents a comparative analysis of the performance metrics for the three models, CCNN, NB, and KNN. The CCNN model has the highest accuracy (99.93%), precision (99.85%), recall (99.99%), and F1-score (99.93%) of all of the examined metrics. On the other hand, the NB categorized documents with a reasonably high recall (84.69%), indicating a substantial number of false positives, but a very poor accuracy (4.19%) and F1-score (7.97%). The KNN model's accuracy, precision, recall, and F1-score are 97.15%, 82.85%, and 93.75%, respectively, making it a rather good model.

TABLE III. COMPARISON OF PERFORMANCE METRICS ON CCFD DATASET

Performance Measures	CCNN	NB[34]	KNN[35]
Accuracy	99.93	97.78	97.15
Precision	99.85	04.19	82.85
Recall	99.99	84.69	93.75
F1-score	99.93	07.97	96.80

The suggested CCNN shows better results on CCFD dataset, with the result of 99.93% being significantly higher compared to traditional tools of classification like NB, KNN. The major strength of the CCNN model is that it can learn and extract deep hierarchical representation of the input data automatically and thus it can better detect the fraud. The combination of such high performance with a lower rate of FP and FN shows that the CCNN is a realistic and scalable option in the realm of financial security and that it is successful in real-time credit card fraud detection applications.

V. CONCLUSION AND FUTURE WORK

Fraud detection in monetary exchanges emphasizes the relevance of modern technology and data analytics as a significant tool in the war on fraud. Financial institutions have deployed AI and ML models and algorithms to identify suspicious activities and anomalous trend that could be an indicator of fraudulent transactions. This study demonstrates how a CCNN model may be used to identify credit card fraud on a large scale and effectively. According to the experimental data, the CCNN model outperforms other conventional ML algorithms like KNN and NB, producing exceptional performance measures with an accuracy of 99.93%. The model with its high recall and precision scores indicates that it is possible to limit FP and FN and to rely on it as the basis to detect the fraud in the financial system in real-time. The confusion matrix and graphical analysis allow us to confirm once again the high learning and generalization power of the model.

Future research will focus on deploying the CCNN model in real-time systems to evaluate its efficiency and scalability. Another top goal is improving model interpretability with explainable AI methods like SHAP or LIME. Additionally, exploring hybrid models, testing on diverse datasets, and improving resilience against adversarial attacks will help increase robustness. Lightweight model versions for mobile and edge devices will support broader, real-world adoption.

REFERENCES

[1] D. Roy and S. Lohana, "Bank Frauds in India: Trends, Modus Operandi and Preventive Measures," 2024.

[2] 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, 2025.

[3] P. Chatterjee, "Proactive Infrastructure Reliability: AI-Powered Predictive Maintenance for Financial Ecosystem Resilience," *J. Artif. Intell. Gen. Sci. ISSN 3006-4023*, vol. 7, no. 01, pp. 291–303, 2024.

[4] M. Shah, P. Shah, and S. Patil, "Secure and Efficient Fraud Detection Using Federated Learning and Distributed Search Databases," in *2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC)*, 2025, pp. 1–6. doi: 10.1109/ICAIC63015.2025.10849280.

[5] A. Alsulami and R. Alabdan, "Fraud Detection in Financial Transactions," *Adv. Appl. Stat.*, vol. 91, no. 8, pp. 969–986, 2024, doi: 10.17654/0972361724052.

[6] B. Chaudhari and S. Chitraju, "Achieving High-Speed Data Consistency in Financial Microservices Platforms Using NoSQL Using Nosql (Mongodb, Redis) Technologies," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 2, pp. 750–759, Jun. 2024, doi: 10.48175/IJARSCT-18890.

[7] V. Pillai, "Anomaly Detection Device for Financial and Insurance Data," *J. AI-Assisted Sci. Discov.*, vol. 4, no. 2, pp. 144–183, 2024.

[8] S. S. Sulaiman, I. Nadher, and S. M. Hameed, "Credit Card Fraud Detection Challenges and Solutions: A Review," *Iraqi J. Sci.*, vol. 65, no. 4, pp. 2287–2303, 2024, doi: 10.24996/ijs.2024.65.4.42.

[9] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[10] B. Ramanujam, "Statistical in Sights in to Anti-Money Laundering: Analyzing Large-Scale Financial Transactions," *Int. J. Eng. Res. Technol.*, vol. 14, no. 4, 2025, doi: 10.17577/IJERTV14IS040136.

[11] 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.

[12] S. Tyagi, "Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions," *Am. Int. J. Bus. Manag.*, vol. 5, no. 01, pp. 5–19, 2022.

[13] M. Nweze, E. K. Avickson, and G. Ekechukwu, "International Journal of Research Publication and Reviews The Role of AI and Machine Learning in Fraud Detection : Enhancing Risk Management in Corporate Finance," vol. 5, no. 10, pp. 2812–2830, 2024.

[14] A. Balasubramanian, "Proactive Machine Learning Approach to Combat Money Laundering in Financial Sectors," *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 2, pp. 1–15, 2019.

[15] G. Mantha, "Transforming the Insurance Industry with Salesforce: Enhancing Customer Engagement and Operational Efficiency," *North Am. J. Eng. Res.*, vol. 5, no. 3, 2024.

[16] 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.

[17] J. K. Chaudhary, S. Tyagi, H. P. Sharma, S. V. Akram, D. R. Sisodia, and D. Kapila, "Machine Learning Model-Based Financial Market Sentiment Prediction and Application," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, May 2023, pp. 1456–1459. doi: 10.1109/ICACITE57410.2023.10183344.

[18] O. I. Odufisan, O. V. Abhulimen, and E. O. Ogunti, "Harnessing artificial intelligence and machine learning for fraud detection and prevention in Nigeria," *J. Econ. Criminol.*, vol. 7, no. October 2024, p. 100127, 2025, doi: 10.1016/j.jeconc.2025.100127.

[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, 2025, doi: 10.38124/ijisrt/25apr1813.

[20] S. Kabade and A. Sharma, "Intelligent Automation in Pension Service Purchases with AI and Cloud Integration for Operational Excellence," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 725–735, Dec. 2024, doi: 10.48175/IJARSCT-14100J.

[21] P. Chatterjee and A. Das, "Adaptive Financial Recommendation Systems Using Generative AI and Multimodal Data," *J. Knowl. Learn. Sci. Technol.*, vol. 4, no. 1, pp. 112–120, 2025.

[22] 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.

[23] V. Pillai, "System and Method for Intelligent Detection and Notification of Anomalies in Financial and Insurance Data using Machine Learning," 202421099024, 2025

[24] S. P. M Shah, "AI/ML Techniques for Real-Time Fraud Detection," *DZone*, 2025.

[25] S. Wawge, "Evaluating Machine Learning and Deep Learning Models for Housing Price Prediction," *IJARSCT*, vol. 5, no. 11, pp. 367–377, 2025.

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

[27] S. S. Nair, G. Lakshmikanthan, N. Belagalla, S. Belagalla, S. K. Ahmad, and S. A. Farooqi, "Leveraging AI and Machine Learning for Enhanced Fraud Detection in Digital Banking System: A Comparative Study," in *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, 2025, pp. 1278–1282. doi: 10.1109/CE2CT64011.2025.10939756.

[28] B. Dharma and D. Latha, "Fraud Detection in Credit Card Transactional Data Using Hybrid Machine Learning Algorithm," in *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, 2025, pp. 213–218. doi: 10.1109/ICMCSI64620.2025.10883549.

[29] A. Singh, K. S. Gill, M. Kumar, and R. Rawat, "Beyond Traditional Methods: Evaluating Advanced Machine Learning Models for Superior Fraud Detection," in *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, 2024, pp. 297–300. doi: 10.1109/ICUIS64676.2024.10866102.

[30] Y. K. Jain, C. A. D. S. Rathore, A. Johrawanshi, A. Maheshwari, A. Pandey, and N. Saxena, "Machine Learning Approaches for Identifying Fraudulent Banking Transactions: A Financial Management Perspective," in *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2024, pp. 1903–1909. doi: 10.1109/ICTACS62700.2024.10841041.

[31] P. Silvia, Q. Aini, E. A. Nabila, Henderi, and H. Nusantoro, "The Role of User Behavior Patterns in Enhancing Fraud Detection in Online Banking: A Bibliometric Analysis," in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2024, pp. 1–6. doi: 10.1109/ICTIIA61827.2024.10761930.

[32] M. A. Islam, A. Nag, S. Chowdhury, S. F. A. Fahim, A. Ghosh, and N. Mumtaj, "Utilization of Encoding, Early Stopping, Hyper Parameter Tuning, and Machine Learning Models for Bank Fraud Detection," in *2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, IEEE, Nov. 2023, pp. 321–327. doi: 10.1109/WIECON-ECE60392.2023.10456503.

[33] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine

learning techniques: a literature review," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, pp. 1–22, 2024, doi: 10.1057/s41599-024-03606-0.

[34] K. Kowsalya, Mrs. Vasumathi, and Dr. S. Selvakani, "Credit Card Fraud Detection Using Machine Learning Algorithm," *EPRA Int. J. Multidiscip. Res.*, pp. 109–116, Mar. 2024, doi: 10.36713/epra16045.

[35] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 International Conference on Computing Networking and Informatics (ICCNI)*, IEEE, Oct. 2017, pp. 1–9. doi: 10.1109/ICCNI.2017.8123782.