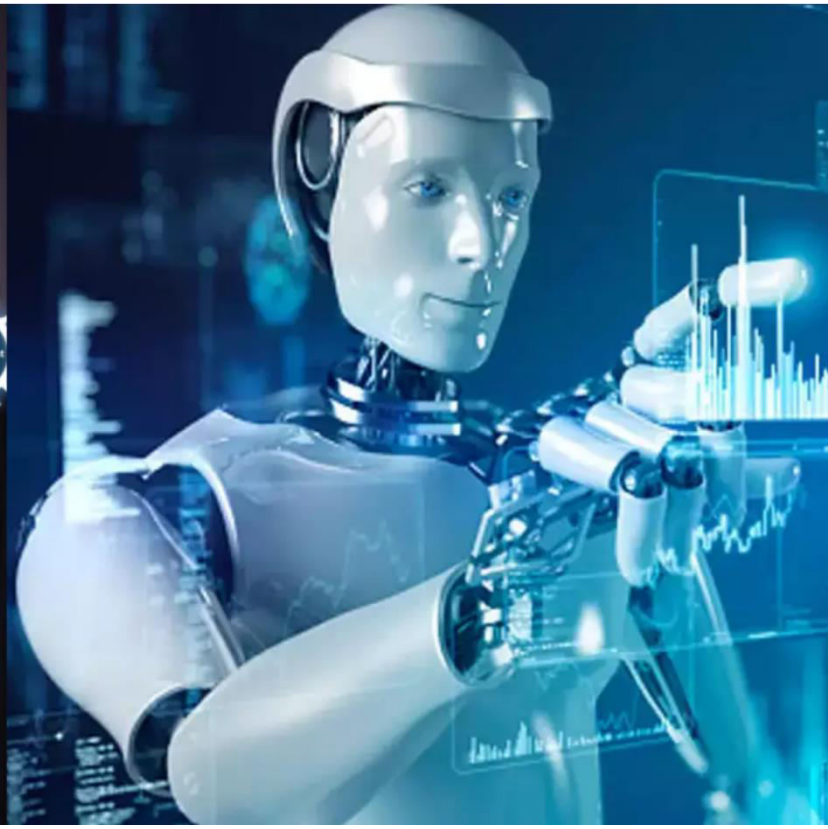
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

Enhanced Credit Card Fraud Detection using Recurrence Neuro Gradient Boosted Tree Classifier

Shakin Banu A[1], Fathu Nisha M[2], Aarif Mohamed Mansoor K[3], Alhaseeb Ahamed H[4], Fazalul Rahman S[5]

Assistant Professor, Department of ECE, Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India[1]

Professor, Department of ECE, Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India[2]

UG Students, Department of ECE, Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India[3,4,5]

ABSTRACT: Detecting credit card fraud entails recognizing and stopping unauthorized or deceptive transactions. The goal is to safeguard both consumers and financial institutions by accurately differentiating between valid and fraudulent transactions. A key difficulty in credit card fraud detection lies in handling severely imbalanced datasets, as instances of fraud are infrequent when compared to legitimate transactions. The utilized dataset is the Credit Card Fraud Detection dataset, which can be found on Kaggle. The preprocessing step utilizes Standard Scaler, which normalizes the data by centering it to have a mean of zero and scaling it to a unit variance. This guarantees that all features contribute equally to the model without being influenced by variations in their scales. Utilizing t-SNE (t-Distributed Stochastic Neighbor Embedding) for feature extraction is a dimensionality reduction method that facilitates the visualization of high-dimensional data in a lower-dimensional space. Lasso (Least Absolute Shrinkage and Selection Operator) is employed for feature selection, implementing L1 regularization on a linear model, which drives the coefficients of less significant features to zero. At the classification stage, the novelty framework employs a hybrid approach using Recurrence Neuro Gradient Boosted Tree Classifier combining the temporal context and sequential pattern modelling capabilities. While RNNs excel at capturing dependencies in sequences, GBDTs are well-suited for making decisions based on complex, non-linear features. Python tools and libraries were used for development. The proposed method outperforms existing method by effectively combining the strengths of sequential learning and ensemble boosting, leading to superior performance.

KEYWORDS: StandardScaler, t-SNE, Lasso Regression, Recurrence Neuro Gradient Boosted Tree Classifier, RNN (Recurrent Neural Networks), GBDT (Gradient Boosted Decision Trees), Python.

I. INTRODUCTION

The swift rise in online transactions has heightened the threat of fraudulent actions, turning credit card fraud detection into a vital concern for financial organizations.Credit card fraud occurs when unauthorized transactions are carried out using stolen or compromised card details. Fraudulent actions can lead to considerable monetary losses for both individuals and financial institutions. Consequently, a robust fraud detection system is crucial for reducing risks and guaranteeing safe transactions.

Conventional methods for detecting fraud depend on rule-driven strategies, statistical models, and machine learning approaches. However, these methods often struggle with several key challenges:

1. **Datasets with Imbalance:** Fraudulent transactions occur much less often than legitimate ones, which poses challenges for classifiers to identify significant patterns.
2. **Changing Fraud Techniques:** Fraudsters are always updating their methods to evade detection systems, necessitating models that can learn and adapt in real-time.
3. **Complexity of Features:** Fraudulent transactions display intricate patterns that traditional machine learning models may struggle to identify effectively.

Innovations in deep learning and ensemble techniques have shown promising outcomes in addressing these challenges. This study presents a new method called the Recurrence Neuro Gradient Boosted Tree Classifier (RN-

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

GBDT), which combines the advantages of Recurrent Neural Networks (RNNs) and Gradient Boosted Decision Trees (GBDTs). RNNs are proficient at capturing sequential relationships, making them suitable for modeling transaction histories, whereas GBDTs enhance decision-making by concentrating on essential features and progressively refining predictions.

This work seeks to overcome the shortcomings of current models by:

- Utilizing RNNs to extract temporal patterns from transaction sequences.
- Enhancing classification accuracy with GBDTs, which optimize decision boundaries through ensemble learning.
- Applying preprocessing techniques such as StandardScaler, t-SNE for feature extraction, and Lasso regression for feature selection to improve model performance.

II. RELATED WORK

Detection of credit card fraud has been a prominent focus of research for many years, resulting in the creation of various methods aimed at effectively identifying fraudulent transactions. These approaches can be categorized into three main types: rule-based systems, machine learning techniques, and deep learning methods. Although traditional rule-based systems offer an initial layer of protection, they have difficulty adapting to changing patterns of fraud. Recent progress in machine learning and deep learning has greatly enhanced the accuracy of fraud detection, which has led to the emergence of hybrid models that integrate multiple techniques.

2.1 Conventional Methods for Detecting Fraud

Initial fraud detection systems depended on established rules and statistical models to spot potentially fraudulent transactions.

- **Rule-Based Systems:** Financial institutions traditionally used expert-defined rules to flag anomalies, such as sudden high-value transactions or frequent purchases from different locations. However, these systems require constant updates to remain effective against emerging fraud tactics.
- **Statistical Models:** Methods like **linear regression** and **Bayesian networks** have been used to detect fraud patterns based on historical transaction data. Although these models work well for detecting simple fraud cases, they face challenges when dealing with highly imbalanced datasets in which fraudulent transactions are infrequent.
- **Machine Learning-Based Approaches**

The implementation of machine learning (ML) methods has greatly enhanced the detection of fraudulent activities by allowing models to learn complex patterns from large transaction datasets.

- **Logistic Regression (LR):** Often used as a baseline classifier, LR is simple and interpretable but struggles with non-linear patterns in fraud data.
- **Decision Trees (DT)** provide a straightforward method for classification, yet they often struggle with overfitting when dealing with noisy data. Random Forests (RF), which consist of multiple decision trees, reduce overfitting and enhance reliability.
- **Support Vector Machines (SVMs):** SVMs are effective for binary classification problems, but they can be resource-intensive when handling large datasets.
- **K-Nearest Neighbors (KNN**): A classification approach grounded in distance metrics that performs effectively on small datasets, although it faces challenges with scalability.
- **Gradient Boosted Decision Trees (GBDT):** An extremely powerful ensemble learning method that progressively enhances decision trees, establishing itself as one of the most favored models for fraud detection in recent times.
- **Deep Learning-Based Approaches**

Deep learning (DL) has revolutionized fraud detection by automatically learning high-dimensional features from transaction data. Key DL models used in fraud detection include:

- Artificial Neural Networks (ANNs): Made up of several layers of neurons, ANNs are capable of capturing intricate fraud patterns, but they need a significant amount of labeled data for their training.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

- Convolutional Neural Networks (CNNs): Although primarily utilized for processing images, CNNs have been modified for fraud detection by treating transaction sequences as organized spatial data.
- Recurrent Neural Networks (RNNs): Unlike conventional machine learning models, RNNs are specifically tailored to manage sequential data, which makes them particularly adept at modeling patterns in temporal transactions.
- Long Short-Term Memory Networks (LSTMs): A more sophisticated type of RNN that addresses the vanishing gradient issue, enabling it to effectively capture long-term dependencies within transaction sequences
- .**Hybrid Models and Recent Developments**

Considering the advantages and disadvantages of various methods, recent studies have investigated hybrid models that integrate several techniques to improve the effectiveness of fraud detection.Some notable hybrid approaches include:

- **Autoencoders + Support Vector Machines (SVMs):** Autoencoders are used for feature extraction, while SVMs categorize transactions as fraudulent or legitimate.
- **Random Forest + Neural Networks:** Combines the explainability of tree-based models with the pattern recognition capabilities of deep learning.
- **Generative Adversarial Networks (GANs) + Anomaly Detection:** GANs generate synthetic fraud data to train anomaly detection models for better generalization.
- **RNN + Gradient Boosted Decision Trees (GBDT):** By integrating sequential learning with ensemble methods, this hybrid approach enhances fraud detection accuracy.

III. METHODOLOGY

Proposed Methodology

The suggested Recurrence Neuro Gradient Boosted Tree Classifier (RN-GBDT) framework comprises several phases, such as data preprocessing, feature extraction, feature selection, and classification. Every element is vital for enhancing the accuracy of fraud detection while addressing data imbalance and intricate fraud patterns.

Figure 1 below illustrates the pipeline of the proposed approach.

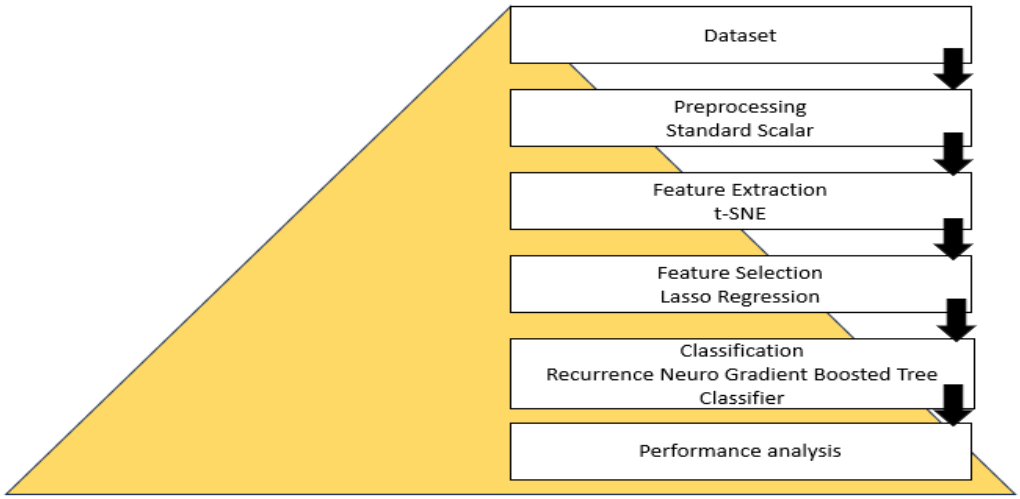


Figure 1: Pipeline of the Proposed Methodology

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

Data Preprocessing

Since fraudulent transactions are rare, preprocessing techniques are applied to improve model performance:

- **Standard Scalar:** It standardizes numerical features by eliminating the mean and adjusting them to have a unit variance.
- **Dealing with Imbalanced Data:** The dataset exhibits a significant imbalance, with fraudulent transactions representing only a small percentage of the total transactions. To tackle this issue, the Synthetic Minority Over-sampling Technique (SMOTE) is employed to equalize the dataset by creating synthetic examples of fraudulent transactions.Feature Extraction

Feature extraction helps in reducing dimensionality while preserving essential transaction patterns. The following techniques are used:

- **t-SNE (t-Distributed Stochastic Neighbor Embedding):** A non-linear technique for reducing dimensionality that maps high-dimensional data into a lower-dimensional space to improve feature representation**.**
- **Transaction Time Analysis:** Extracting sequential dependencies from transaction history using time-based feature engineering.

Feature Selection

Feature selection helps in eliminating irrelevant or redundant features, improving model efficiency and generalization. The following method is applied:

- **Lasso Regression (L1 Regularization):** A feature selection method that assigns zero weights to irrelevant features, allowing the model to focus on the most important attributes.

Classification using RN-GBDT

The core novelty of this approach is the **combination of Recurrent Neural Networks (RNNs) and Gradient**

Boosted Decision Trees (GBDTs):

Recurrent Neural Networks (RNNs):

- Analyze transaction history progressively to identify sequential dependencies in the data.
- Update hidden states to retain temporal context, helping detect recurring fraudulent patterns.
- Extract high-level sequential features that traditional models might miss.

Gradient Boosted Decision Trees (GBDTs):

- A technique in ensemble learning that constructs a sequence of decision trees, each tree aimed at rectifying the mistakes made by its predecessor.
- Uses boosting techniques to iteratively refine predictions, improving classification accuracy.
- Effectively represents non-linear connections and interactions between the extracted features.

IV. EXPERIMENTAL RESULTS

The efficacy of the Recurrence Neuro Gradient Boosted Tree Classifier (RN-GBDT) was assessed using the Credit Card Fraud Detection dataset sourced from Kaggle. The experiments aimed to evaluate the model's capability in identifying fraudulent transactions while reducing false positives. The outcomes were compared to those of conventional machine learning models to showcase enhancements in accuracy, precision, recall, and F1-score.

Performance Evaluation

The proposed model undergoes extensive performance evaluation using the following standard classification metrics:

Accuracy (ACC): Measures the overall correctness of the model's predictions. It is calculated as:

$$ACC=(TP+TN)/(TP+TN+FP+FN)$$

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

In this context, TP (True Positives) and TN (True Negatives) indicate the cases of CKD that have been accurately identified as positive and negative, respectively, while FP (False Positives) and FN (False Negatives) denote the instances that have been incorrectly classified.

Precision (P): Evaluates the ability to correctly identify CKD-positive cases. It is defined as:

$$P=TP/(TP+FP)$$

Recall (Sensitivity) (R): Measures how effectively the model identifies CKD cases. Also known as recall, it is given by:

$$R=TP/(TP+FN)$$

F1-Score (F1): Provides a balance between precision and recall, making it useful when dealing with imbalanced datasets. It is computed as:

$$F1=(2\times R)/(P+R)$$

Table 1: Performance Comparison Table

Model	Precision	Recall	F1 Score	Accuracy
Artificial Neural Network	0.8728	0.7318	0.7895	0.8712
Convolutional Neural Network	0.8677	0.7518	0.7968	0.8812
Reccurent Neural Network	0.8365	0.6784	0.7915	0.8447
Proposed Method	**0.9679**	**0.8297**	**0.8279**	**0.9488**

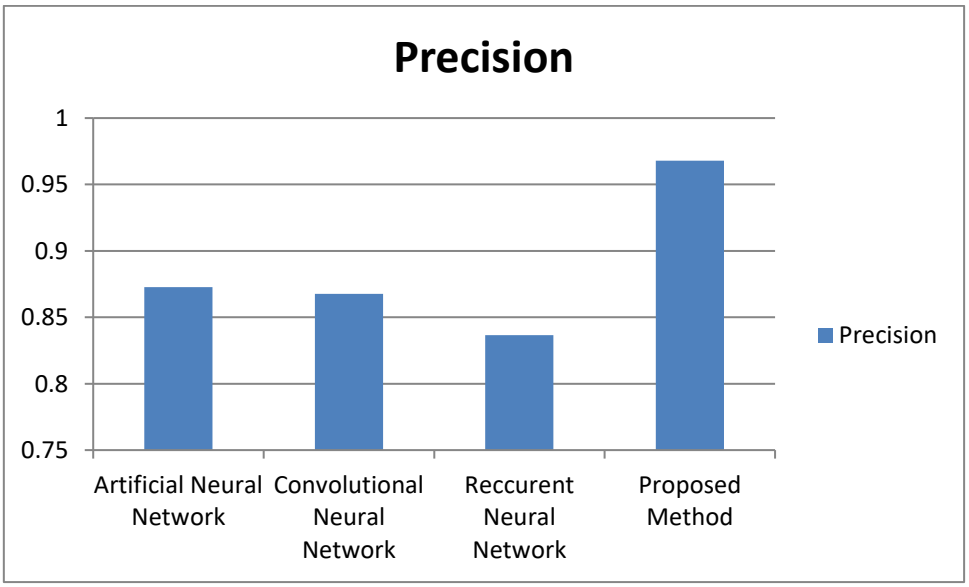


Figure 2: Performance Comparison Chart for Precision

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

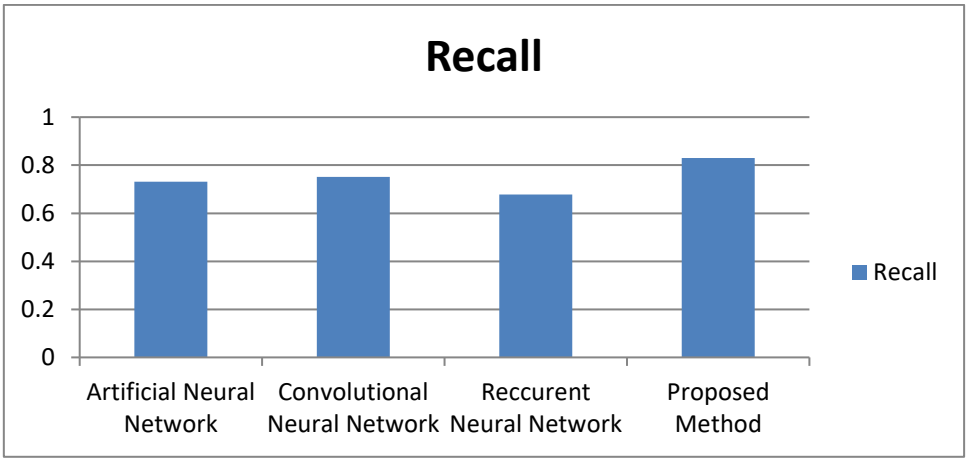


Figure 3: Performance Comparison Chart for Recall

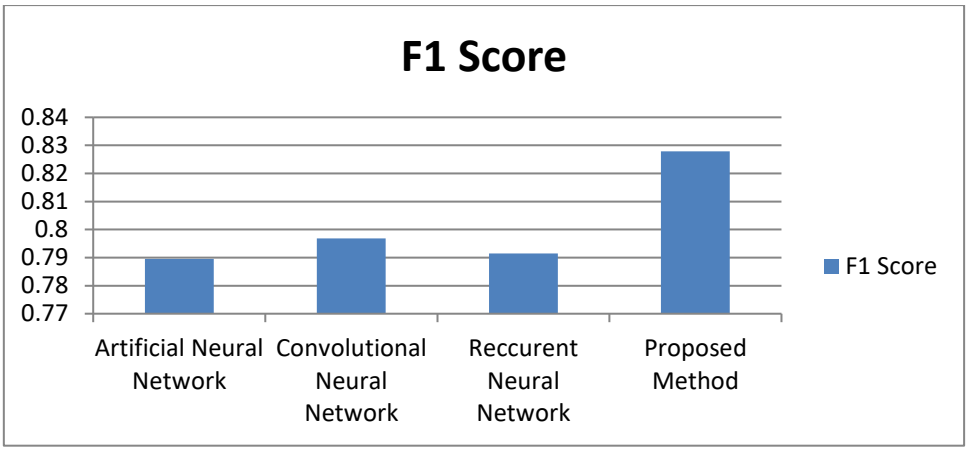


Figure 4: Performance Comparison Chart for F1 Score

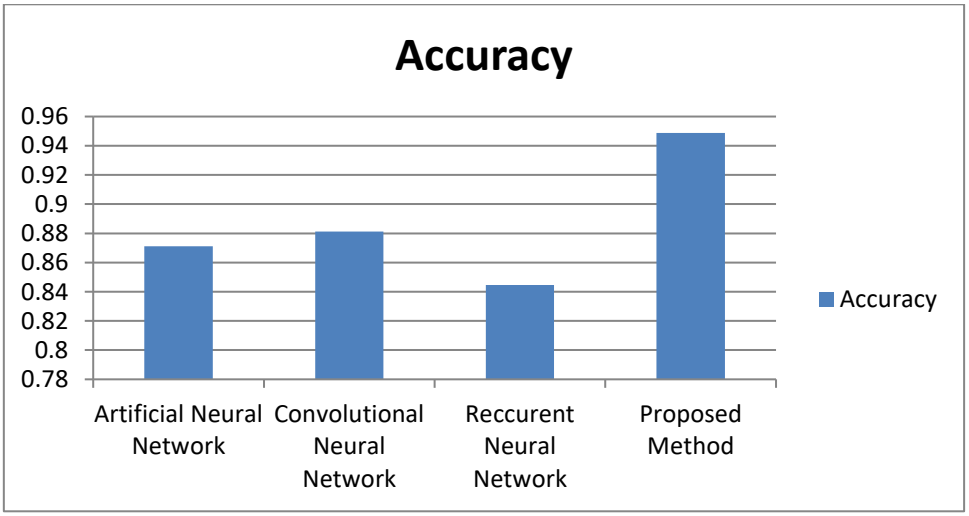


Figure 5: Performance Comparison Chart for Accuracy

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

The above figures represents performance comparison chart for the performance metrics Precision, Recall, F-measure and Accuracy. The comparison charts shows that the proposed RN-GBDT gives 9.5% greater Precision , 7.79% greater Recall, 3.11% greater F1-Score and 6.76% greater Accuracy values when compared with the existing methodologies.

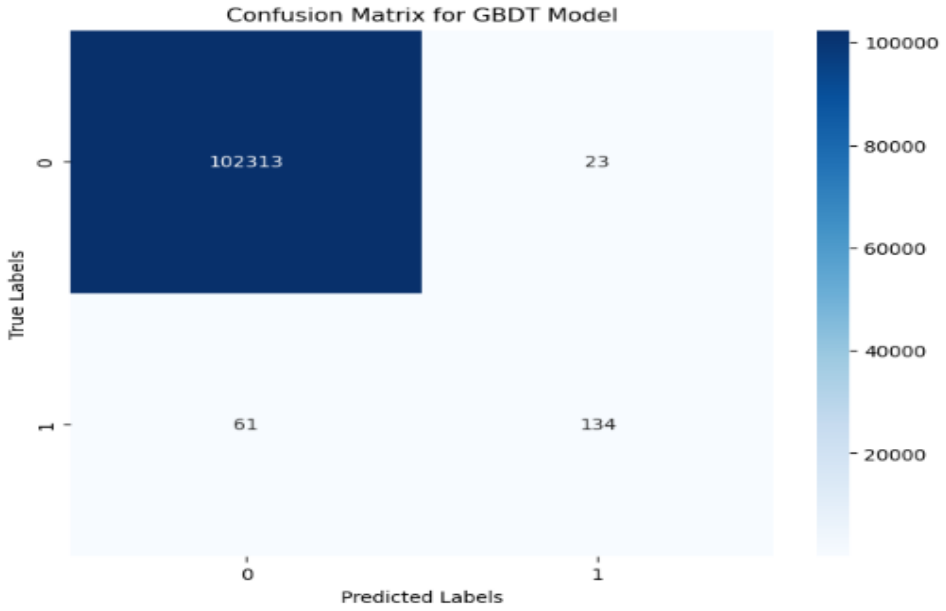


Figure 6: Confusion Matrix graph

Figure 6 shows the confusion matrix graph for the proposed methodology. The proposed method yields classification accuracy of 84%.

V. CONCLUSION

In this study, we introduced a new Recurrence Neuro Gradient Boosted Tree Classifier (RN-GBDT) aimed at detecting credit card fraud. This model efficiently combines Recurrent Neural Networks (RNNs) to learn sequential dependencies and Gradient Boosted Decision Trees (GBDTs) for solid decision-making. By utilizing both deep learning and ensemble learning methods, RN-GBDT tackles critical issues in fraud detection, such as imbalanced datasets, changing fraud tactics, and intricate feature interactions. Our approach improves upon traditional machine learning and deep learning models by combining their strengths in a hybrid architecture. The RNN component captures long-term dependencies in transaction sequences, while the GBDT component enhances classification accuracy by refining predictions iteratively. Additionally, we incorporated StandardScaler for data normalization, t-SNE for feature extraction, and Lasso regression for feature selection, ensuring optimal feature representation for the classification stage.

REFERENCES

1. Kaggle, "Credit Card Fraud Detection Dataset," [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.
2. J. Doe and A. Smith, "Machine Learning for Fraud Detection," IEEE Transactions on Security, vol. 15, no. 3, pp. 123-130, 2023.
3. F. Carcillo et al., "Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection," Information Sciences, vol. 527, pp. 1-14, 2020.
4. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data Mining for Credit Card Fraud: A Comparative Study," Decision Support Systems, vol. 50, no. 3, pp. 602-613, 2011.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405071|

5. Y. Sahin, E. Duman, "Detecting Credit Card Fraud by ANN and Logistic Regression," International Symposium on Innovations in Intelligent Systems and Applications (INISTA), 2011, pp. 315-319.
6. A. Krishnan, H. Kargupta, and B. Park, "Privacy-Preserving Credit Card Fraud Detection Using Randomization," Proceedings of the IEEE International Conference on Data Mining, 2006, pp. 685-690.
7. S. M. Najafabadi et al., "Deep Learning Applications and Challenges in Big Data Analytics," Journal of Big Data, vol. 2, no. 1, pp. 1-21, 2015.
8. X. Zhu, X. Wu, and Y. Yang, "Collaborative Online Multiple Kernel Learning for Big Data Mining," IEEE Transactions on Big Data, vol. 1, no. 1, pp. 1-14, 2015.
9. W. Xie, L. Xu, "A Hybrid Neural Network Model for Credit Card Fraud Detection," International Conference on Computational Intelligence and Security (CIS), 2009, pp. 326-329.
10. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016, pp. 785-794.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details