A Hybrid Deep Learning Model For Detecting Credit Card Fraud Using CNN–BiLSTM With An Attention Mechanism And Focal Loss Optimization

A.Keerthi[1], P.Devalekka[2], M.Sahana[3],DrR.Punithavathi[4]
[1, 2, 3] Dept of Artificial Intelligence and Data Science
[4]HOD, Dept of Artificial Intelligence and Data Science
[1, 2, 3, 4] Chettinad College of Engineering and Technology,Karur,India

Abstract- *Detecting credit card fraud is a complex task due to the significant imbalance in data and the constantly changing nature of fraudulent activities. This research introduces a hybrid deep learning model that combines CNN, BiLSTM, and an attention mechanism to effectively identify spatial and temporal patterns in transactions. To maintain regional specificity, separate models were developed for datasets from India and Europe. To tackle class imbalance without introducing synthetic bias, Focal Loss with adaptive class weighting was employed. The experiments revealed that the model for the European dataset achieved an accuracy of 99.32% and an F1-score of 97.29%, while the model for the Indian dataset reached an accuracy of 98.67% and an F1-score of 95.80%. The use of attention mechanisms enhanced the relevance of features and overall performance, and SHAP-based explainability improved the interpretability of the model. The system was deployed using Stream lit for real-time fraud detection, offering a scalable and precise solution tailored to region-specific financial fraud detection.*

Keywords: Credit Card Fraud Detection,Deep Learning, Attention-Based Models, Focal Loss

I. INTRODUCTION

The swift expansion of digital payment platforms and online financial transactions has greatly heightened the risk of credit card fraud, making fraud detection a major issue for financial institutions globally [1, 4, 9]. With millions of transactions taking place each day, fraudulent activities are often concealed within highly imbalanced datasets, where legitimate transactions far exceed fraudulent ones [2, 5, 10]. This imbalance, along with the constantly changing nature of fraud tactics, renders traditional rule-based and statistical methods inadequate for precise and prompt detection [3, 6, 14].

In recent times, machine learning and deep learning methods have demonstrated encouraging outcomes in tackling fraud detection issues [7, 11, 12]. Approaches like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are frequently employed to identify intricate patterns within transactional data [8, 13]. Nonetheless, numerous current models struggle to simultaneously capture both spatial and temporal dependencies and often neglect regional differences in transaction behavior [9, 11]. Furthermore, typical strategies for managing class imbalance, such as oversampling, can introduce synthetic bias and diminish the model's ability to generalize [5, 6].

This study introduces a hybrid deep learning framework designed to address existing limitations by combining CNN, Bidirectional LSTM (BiLSTM), and an attention mechanism to boost feature extraction and enhance prediction accuracy [8, 12]. A significant element of this research is the use of dataset-specific learning, where distinct models are developed for Indian and European transaction datasets. Each model is dedicated to its respective dataset, ensuring that the unique transaction patterns of each region are accurately captured and learned [2, 7]. Additionally, Focal Loss with adaptive class weighting is utilized to effectively manage class imbalance without the need for synthetic data generation [5, 9].

The proposed system undergoes evaluation using real-world datasets and is further improved with SHAP-based explainability to ensure transparency in its predictions [1, 3]. Moreover, the model is implemented as a real-time application through Streamlit, facilitating interactive fraud detection. This work ultimately seeks to offer a scalable, precise, and interpretable solution for detecting credit card fraud, thereby enhancing the security and reliability of financial systems [6, 10, 14].

II. LITERATURE SURVEY

Numerous strategies have been employed to detect credit card fraud, including conventional machine learning techniques like logistic regression, decision trees, and random

forests. These methods depend on feature engineering and data preprocessing to uncover patterns in transaction data, with genetic algorithms often used to enhance feature selection. Nevertheless, because fraud datasets are highly imbalanced, these models frequently exhibit a bias toward legitimate transactions, resulting in subpar fraud detection. To enhance performance, ensemble and hybrid methods have been developed, which combine multiple models or utilize clustering techniques to boost accuracy and decrease false positives. Despite these advancements, these methods still rely on manual feature engineering and lack interpretability, posing challenges for their application in real-world banking systems.

Deep learning methods have enhanced fraud detection by identifying intricate patterns and the behavior of sequential transactions. Models like LSTM and autoencoders excel at understanding temporal dependencies and spotting anomalies more precisely than traditional techniques. Nonetheless, these models encounter difficulties such as managing extreme class imbalance and offering clear explanations for their predictions. Consequently, there is a demand for a sophisticated approach that combines feature extraction, sequential learning, imbalance management, and interpretability, which is addressed by the proposed hybrid CNN–BiLSTM–Attention model with Focal Loss.

III. METHODOLOGY

A. System Architecture Overview

The proposed system presents a deep learning-based framework for credit card fraud detection designed to address class imbalance and capture sequential patterns in transaction data. The architecture follows a structured pipeline consisting of five stages: data ingestion, preprocessing and feature engineering, hybrid deep learning model development, model training using Focal Loss, and explainability with deployment. Transaction data from both European and Indian datasets is collected in CSV format and processed through normalization, encoding of categorical variables, missing value handling, and train-test splitting. The processed data is then passed into a hybrid deep learning model for feature learning and classification as shown in the Fig 3.1;

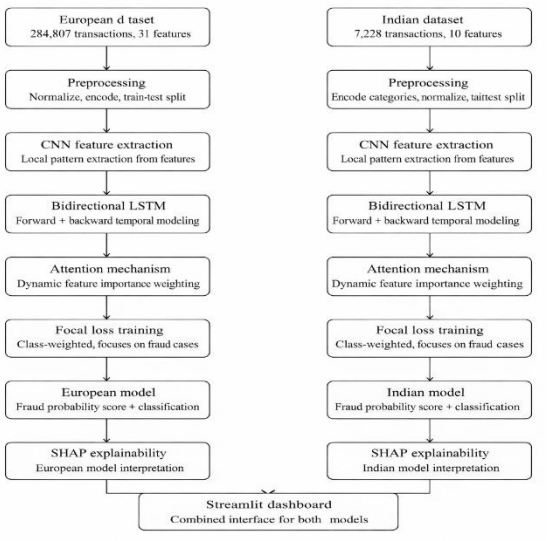

Fig: 3.1 Workflow Diagram

B. Data Collection and Preprocessing

The system utilizes two datasets to ensure robustness and generalization. The European dataset contains 284,807 transactions with a highly imbalanced distribution of fraud cases (approximately 0.17%) as shown in Fig 3.2, where features are anonymized using PCA transformation along with Time and Amount attributes. The Indian dataset contains around 7,000 transactions and includes categorical features such as location, transaction type, and card category. The Dataset has 68.66 legit and 31.34 fraud transaction as shown in the Fig 3.3. Preprocessing involves handling missing values, feature scaling using standardization, encoding categorical variables using one-hot encoding, and splitting the dataset into training and testing sets in an 80:20 ratio. These steps ensure clean and model-ready input data.

Fig: 3.2European Dataset Fraud vs Legit Ratio

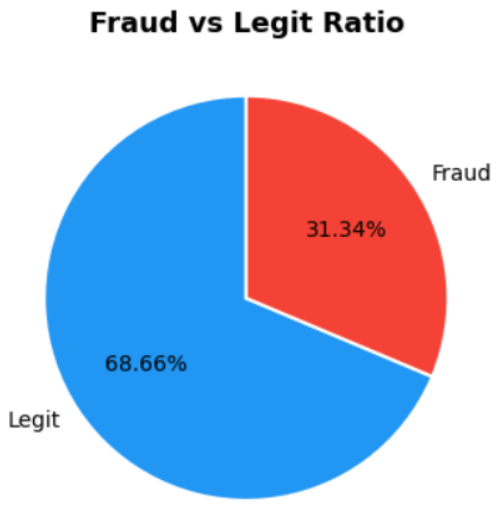

Fig: 3.3Indian Dataset Fraud vs Legit Ratio

C. Feature Extraction and Model Design

A hybrid deep learning architecture is proposed to capture both spatial and temporal patterns in transaction data. The Convolutional Neural Network (CNN) layer extracts local feature relationships and reduces noise by learning high-level feature representations. The Bidirectional Long Short-Term Memory (BiLSTM) layer processes transaction sequences in both forward and backward directions, enabling the model to learn long-term dependencies in user behavior. An Attention mechanism is incorporated to dynamically assign weights to important features, improving both performance and interpretability. The final dense layers perform feature transformation and output a fraud probability score using a sigmoid activation function, where transactions are classified based on a predefined threshold.

D. Model Training and Evaluation

To handle severe class imbalance, the model is trained using Focal Loss, which focuses more on difficult-to-classify fraud cases while reducing the influence of majority class samples. This improves the model's ability to detect rare fraudulent transactions. The model is optimized using adaptive optimizers and trained with techniques such as batch processing, early stopping, and class weighting to prevent overfitting. Performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Among these, recall is considered the most critical metric since minimizing false negatives is essential in fraud detection systems.

E. Explainability using SHAP

To enhance transparency and trust in predictions, SHAP (SHapley Additive Explanations) is integrated into the system. It provides both local and global interpretability by explaining the contribution of each feature toward a model's decision. This allows financial institutions to understand why a transaction is classified as fraudulent and ensures compliance with regulatory requirements. SHAP also helps in identifying the most influential features affecting fraud detection performance.

F. Deployment using Streamlit Dashboard

The trained model is deployed using a Streamlit-based interactive dashboard for real-time fraud detection. The interface allows users to upload transaction datasets for batch prediction or test individual transactions in real time. It displays fraud probability scores, classification results, and SHAP-based explanations for interpretability. The system loads pre-trained models and preprocessing components during runtime, ensuring fast inference without retraining. This makes the deployment efficient, scalable, and suitable for real-world financial applications.

IV. RESEARCH AND DEVELOPMENT

Credit card fraud detection systems encounter significant challenges with current methodologies. Rule-based systems are inflexible and necessitate ongoing manual updates, which results in high false-positive rates and diminished adaptability to changing fraud patterns. Conventional machine learning models like logistic regression and decision trees are affected by class imbalance bias, as fraudulent cases represent a minor fraction of the dataset, leading to inadequate fraud detection (low recall). To tackle this imbalance, techniques such as SMOTE are frequently employed; however, they introduce synthetic noise by creating unrealistic fraud samples, thereby reducing the model's ability to generalize. Initial deep learning approaches using unidirectional LSTM fail to fully capture temporal dependencies, as they process sequences in only one direction.

Moreover, the majority of models operate as opaque systems, lacking clarity, which hinders their application in practical financial environments that demand openness and adherence to regulations. In addition, numerous studies depend on a single dataset, often from Europe, which limits the ability to apply findings to various regions and transaction patterns.

The proposed system overcomes significant shortcomings of current fraud detection techniques by employing a comprehensive hybrid deep learning framework.

It substitutes inflexible rule-based systems with a model driven by data, which adjusts to changing fraud trends in both Indian and European datasets. To address class imbalance, Focal Loss with adaptive class weights is utilized, enhancing fraud detection by concentrating on difficult-to-classify minority samples and eliminating the need for SMOTE and its related synthetic noise. Temporal dependencies are more effectively captured using a Bidirectional LSTM, which analyses transaction sequences in both forward and backward directions. A CNN layer is incorporated to extract local feature patterns, creating a CNN–BiLSTM–Attention pipeline that captures both spatial and temporal information. Model interpretability is improved through SHAP and attention visualization, ensuring transparency and compliance with regulations. Furthermore, testing on two different datasets shows strong cross-regional generalization and robustness in real-world financial settings.

V. CONCLUSION

To overcome this issue, instead of using traditional oversampling techniques, Focal Loss with class weighting is applied so that the model gives more importance to difficult and rare fraud cases. The proposed hybrid model combining CNN, Bidirectional LSTM, and attention mechanism helps in capturing both feature-level patterns and sequential transaction behaviour. The results show that the model improves fraud detection performance while maintaining reliability. In addition, the use of SHAP values and attention visualization provides better understanding of the model's decisions, making the system more suitable for real-world applications. Finally, the confusion matrix and the accuracy scores are calculated.

REFERENCES

[1] A. K. Verma et al., "Review of Machine Learning Approach on Credit Card Fraud Detection," Discover Artificial Intelligence, Springer Nature, 2022.[Online]. Available: https://doi.org/10.1007/s44230- 022-00004-0

[2] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," Big Data and Cognitive Computing, vol. 8, no. 1, 2024.[Online]. Available: https://doi.org/10.3390/bdcc8010006

[3] A. R. Khalid et al., "Voting Ensemble-Based Credit Card Fraud Detection Using Machine Learning," Big Data and Cognitive Computing, 2024. [Online]. Available: https://www.mdpi.com/2504- 2289/8/1/6

[4] D. Rai and J. S. N., "Credit Card Fraud Detection Using Machine Learning and Data Mining Techniques — A Literature Survey,"Zenodo, 2023.[Online]. Available: https://doi.org/10.5281/zenodo.8190094

[5] E. Ileberi, Y. Sun, and Z. Wang, "A Machine Learning-Based Credit Card Fraud Detection Using Genetic Algorithm for Feature Selection," Journal of Big Data, vol. 9, no. 1, 2022. [Online]. Available: https://doi.org/10.1186/s40537-022-00573-8

[6] J. Wang et al., "Enhancing Credit Card Fraud Detection Using DBSCAN-Augmented Disjunctive Voting Ensemble,"Scientific Reports, vol. 15, 2025. [Online]. Available: https://doi.org/10.1038/s41598-025- 22960-w

[7] L. Zhang et al., "Enhancing Fraud Detection in Credit Card Transactions: A Comparative Study of Machine Learning Models," Computational Economics, 2025.[Online]. Available: https://doi.org/10.1007/s10614-025-11071-3

[8] R. K. Mishra et al., "Autoencoder and LSTM-Based Credit Card Fraud Detection," SN Computer Science, vol. 4, no. 4, 2023. [Online]. Available: https://doi.org/10.1007/s42979-023- 01977-w

[9] M. Al-Hassan et al., "A Systematic Review of AI-Enhanced Techniques in Credit Card Fraud Detection," Journal of Big Data, vol. 12, no. 1, 2025. [Online]. Available: https://doi.org/10.1186/s40537-024- 01048-8

[10] M. A. Gill, M. Qureshi, A. Rasool, and M. M. Hassan, "Detection of Credit Card Fraud Through Machine Learning in Banking Industry," Journal of Computing & Biomedical Informatics, 2021.Online.Available: https://jcbi.org/index.php/Main/article/view/204

[11] R. Sharma and P. K. Singh, "Credit Card Fraud Detection: A Comparative Study of Machine Learning and Deep Learning Methods," Engineering, Technology & Applied Science Research, vol. 10, no. 5, 2023.[Online]. Available: https://doi.org/10.47191/etj/v10i05.45

[12] S. Mehta et al., "A Comparative Study of Machine Learning and Deep Learning Models for Credit Card Fraud Detection," Computational Economics, 2025. [Online]. Available: https://doi.org/10.1007/s10614-025-11071-3

[13] S. R. Patel et al., "Credit Card Fraud Detection Using Deep Learning: A Survey,"DeepAI, 2023.Online. Available: https://deepai.org/publication/credit-card-fraud-detectionusing-machine-learning-a-survey

[14] V. Kumar K. S. et al., "Credit Card Fraud Detection Using Machine Learning Algorithms," International Journal of Engineering Research & Technology, 2021.[Online]. Available: https://www.ijert.org/credit-card-fraud-detection-usingmachine-learning-algorithms