Volume 12, No.10, October 2025

Journal of Global Research in Mathematical Archives

ISSN 2320 – 5822

UGC Approved Journal

RESEARCH PAPER

Available online at http://www.jgrma.info

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR DETECT FINANCIAL FRAUDS: IMPROVING TRANSPARENCY AND RUST IN PREDICTIVE MODELS

Amit Jain[1]

[1] Professor, Department of Computer Science and Engineering, OP Jindal University, Raigarh(C.G)
amit.jain@opju.ac.in

Abstract: Financial fraud continues to undermine the stability of economies and institutions worldwide, with billions lost annually despite conventional prevention strategies. The complexities of modern fraud schemes are often beyond the capacity of conventional rule-based systems; hence, more sophisticated approaches are required. This paper uses the IEEE-CIS dataset to offer a methodology for financial fraud detection based on explainable artificial intelligence (XAI). The methodology incorporates data preprocessing, feature encoding, PCA-based dimensionality reduction, and class imbalance handling via random oversampling. An Artificial Neural Network (ANN) model is implemented and compared against Decision Tree, LightGBM, Logistic Regression, and CNN classifiers. Results indicate that while ANN achieves high accuracy (ACC) (97.56%), recal (REC)l remains moderate, reflecting challenges in detecting minority fraud cases. To enhance transparency, interpretability methods such as SHAP and LIME are applied, offering clear insights into model decision-making. A comparative analysis shows that CNN delivers the best overall balance across metrics, while LightGBM demonstrates superior precision. The study helps bridge the gap between predictive performance and interpretability, ensuring reliable and regulation-compliant fraud detection. Additionally, it provides a framework adaptable to diverse financial datasets, enabling future improvements in fraud prevention strategies.

Keywords: Financial Fraud Detection, Explainable Artificial Intelligence (XAI), Machine Learning, Predictive Models, SHAP, LIME.

1 INTRODUCTION

In the digital age, the financial industry is changing dramatically due to the necessity to quickly adjust to new social and economic realities and the speed at which technology is developing[1][2]. A key focus of this transition is sustainable finance, which is crucial to the shift towards a sustainable economic model. Sustainable finance is depending more and more on cutting-edge financial tools created to raise funds for long-term initiatives, including promoting the circular economy, improving energy efficiency, and reducing the effects of climate change.

A financial fraud occurs when a person gains money through deceitful and unlawful ways[3]. The corporate world, banking, insurance, and taxation are just a few of the settings where financial fraud can happen. Recent years have seen an increase in the issues that financial crime causes for sectors and enterprises, such as financial transaction fraud and money laundering[4]. Despite numerous efforts to combat financial fraud, it continues to have a detrimental impact on the economy and society, resulting in significant daily losses. [5]. The term "financial system" refers to the network of institutions, both specialized and non-specialized, services, and products that facilitate the flow of funds[6][7]. This system also includes financial interrelationships and market-adopted procedures and practices. In markets for goods or services, buyers exchange their money for something right away.

Artificial Intelligence (AI) uses its capacity to examine enormous volumes of data and spot trends, and predict fraudulent conduct to propose innovative solutions to this growing issue[8]. This abstract explores the many AI methods and how they are used to combat fraud, emphasizing how they are revolutionizing the security environment[9]. Fraudulent transactions are commonly detected using ML techniques, in particular supervised learning methods that make use of past data to train new models, including neural networks and decision trees[10]. Because these models are able to identify minor patterns that traditional rule-based systems would overlook, they are able to differentiate between transactions that are real and scams.

Machine learning (ML) and advanced analytics are crucial instruments in the battle against fraud. These technologies enable the analysis of multiple data sources, including financial exchanges, personnel files, consumer conduct, and communication systems[11]. They uncover hidden connections and quickly identify questionable behavior [12]. A proactive defense is required because stopping the development of fraud schemes is becoming more and more crucial[13]. These days, even the most basic theories guiding financial regulation need to be reconsidered[14]. Integrating AI into financial institutions' present systems and procedures is one of the largest issues confronting the financial sector at the moment.

1.1 Significance and contribution of this study

The combined challenges of obtaining high fraud detection accuracy and preserving model transparency are the focus of this work. The study closes the gap between interpretability and predictive performance by combining explainable AI techniques with sophisticated ML models, guaranteeing that financial institutions may deploy AI-driven fraud detection systems with increased responsibility and confidence. The study's main contributions are listed below:

- Utilized the IEEE-CIS Fraud Detection dataset, which addresses the very unbalanced class distribution, was obtained from Kaggle and contains about 590,000 transactions.
- Comprehensive preparation was put into place, which included feature encoding, PCA, cleansing the data, dealing with missing values, and using random oversampling to handle class imbalance.
- To identify intricate non-linear patterns in financial transaction data, an ANN model was created.
- Applied **sigmoid activation functions** to add non-linearity to enhance forecasting accuracy.
- The model's performance was correctly evaluated using the following metrics: confusion matrix, F1-score, recall, accuracy, and precision.

1.2 Justification and Novelty

This study is justified by the notable increase in financial fraud incidents and by the inability of traditional detection technologies to keep pace with the intricate, evolving nature of fraud. Rule-based approaches, although widely used, often fail to effectively address the complexities of real-world settings where fraudulent transactions are rare and highly variable. This lack of transparency undermines trust and adoption among financial institutions, regulators, and end users. This study is interesting because it uses the IEEE-CIS dataset to incorporate explainable AI methods into a comprehensive fraud detection pipeline, including SHAP and LIME. Unlike many prior studies that emphasize accuracy alone, this study ensures that forecasts are not only precise but also comprehensible. The combination of preprocessing techniques, class imbalance handling, and dimensionality reduction, along with XAI-based interpretability, provides a robust and transparent solution. This dual emphasis on detection capability and explanation makes the approach unique, balancing performance with accountability and supporting broader adoption in sensitive financial domains.

1.3 Structure of the paper

The paper is structured as follows: Section 2 provides an overview of pertinent research on explainable AI and financial fraud detection. Section 3 describes the suggested approach, including data preparation, model implementation, and interpretability techniques. Section 4 presents the test findings and comparisons. Finally, Section 5 offers a summary and prospects for further study.

2 LITERATURE REVIEW

This survey of the literature focuses on current developments in ML and DL methods for identifying financial fraud. Studies utilize various data inputs and feature engineering to improve fraud identification. Advanced techniques, including ensemble models and DL, have demonstrated enhanced accuracy and reliability, providing more effective tools for monitoring and preventing fraudulent transactions.

Jabeen *et al.* (2025) suggested CNN-LSTM (CLST) model, which has a ROC-AUC of 0.9733, an F1-score of 76% for fraudulent transactions, a recall of 83%, and an accuracy of 70%. Hyperparameter optimization improves the suggested model's performance, yielding an F1-score of 91% for hazardous scenarios, an accuracy of 83%, and a recall of 99%. Furthermore, it achieves an ROC-AUC of 0.9995, This implies nearly flawless fraud detection and a very low incidence of false negatives. To improve fraud detection accuracy and address class imbalance, a hybrid DL model comprising a CNN, an LSTM, and a fully connected output layer is recommended. Spatial characteristics are handled by a CNN, sequential information by an LSTM, and ultimate decision-making is handled by a fully linked output layer[15].

Shah (2025) has employed ML techniques using Kaggle's Financial Fraud Detection Dataset and applying data preprocessing, feature engineering, and class balancing. LightGBM (LGBM), RF, AdaBoost, and a Voting Classifier are among the models that are trained and refined with GridSearchCV to increase ACC. Results Achieved: LGBM achieves the highest ACC (90.20%), followed by the Voting Classifier (90.02%), while RF and AdaBoost record 89.26% and 88.37%, respectively[16].

Singh et al. (2024) described an autoencoder-based fraud detection system that swiftly identifies suspicious bank transactions. The system's experimental study shows that it achieves a REC of 90%, a PRE of 92.3%, and a detection ACC of 98%.5% during the query object detection step. 91% F1, 100% REC, 92.8% PRE, and 95.1% ACC[17].

Geng and Zhang (2023) propose a dual adversarial learning unsupervised network to detect credit card fraud. In contrast to traditional approaches to anomaly identification, this methodology prioritizes the simultaneous investigation of initial and intermediate features. Compared to current fraud detection approaches, the system outperforms them with an MCC of 0.8456%, an accuracy of 0.9224%, and an F1 score of 0.9208%, according to research using the European cardholder dataset. But the frequency

of credit card theft has nonetheless caused incalculable damage to consumers, businesses, and institutions. The main classification techniques used in modern fraud detection procedures include CNN, LSTM, and DNN [18].

The Rallapalli, Hegde, and Thatikonda (2023) The BMOA is used in three steps of preprocessing the dataset to guarantee data balance. Lastly, classification is done during fraud detection using Latent Variable SVM (LV-SVM) and Least Squares SVM (LS-SVM). There has also been a comparison between the suggested approach and current approaches. With a 98% accuracy rate, the suggested method successfully recognized fraudulent transactions, compared to cutting-edge methods, which are clearly superior[19].

Xiuguo and Shengyong (2022) have created a set of financial and non-financial indicators, and then used word vectors to extract those that are contained inside the annual reports of Chinese corporations that are registered on the Stock Exchanges, Medical and analysis part. The empirical results reveal a considerable increase in performance between the suggested DL strategies and traditional ML methods, with testing samples showing 94.98% and 94.62% accuracy rates in classification. This implies that the MD&A section's derived textual characteristics show encouraging classification results and greatly improve the identification of financial fraud[20].

Table 1 summarizes recent studies on ML-based financial fraud detection. Results demonstrate high accuracy using methods such as Random Forest, XGBoost, ensemble models, and DL. Limitations include dataset specificity and scalability, while future work focuses on real-time deployment, broader datasets, and enhanced interpretability.

Table 1: Summary of Previous Work on Summary of Financial Fraud Detection

Authors (Year)	**Dataset**	**Methods Used**	**Key Findings**	**Limitations & Future Work**
Jabeen et al., (2025)	Credit card transactions (European cardholder dataset)	CNN-LSTM (CLST) model with fully connected output; hyperparameter optimization	Initial performance: Recall 83%, Precision 70%, F1-score 76%, ROC-AUC 0.9733. Optimized performance: Recall 99%, Precision 83%, F1-score 91%, ROC-AUC 0.9995. CNN captures spatial features; LSTM captures sequential patterns; addresses class imbalance effectively	High computational cost due to deep hybrid architecture; potential overfitting on imbalanced data; future work could explore lightweight models or XAI integration for interpretability
Shah (2025)	Financial Fraud Detection Dataset (Kaggle)	Random Forest, AdaBoost, LightGBM (LGBM), Voting Classifier, GridSearchCV, SHAP for feature importance	LGBM achieved highest accuracy: 90.20%, Voting Classifier: 90.02%, RF: 89.26%, AdaBoost: 88.37%. SHAP improved interpretability and transparency of results.	Enhance scalability and performance on real-time data streams.
Singh et al. (2024)	Real-time financial transaction data	Autoencoders for anomaly detection	Detection accuracy: 95.1%, precision: 92.8%, recall: 100%, F1-score: 91. Showed strong ability to detect fraudulent activities through learning from successive transactions.	Further testing required on large-scale, diverse datasets.
Geng et.al. (2023)	European cardholder dataset	Unsupervised anomaly detection network using dual adversarial learning	Accuracy 0.9224, F1-score 0.9208, MCC 0.8456; considers both original and intermediate features for better anomaly detection; outperforms conventional methods	Focused on unsupervised learning may miss evolving fraud patterns; lacks explainability; future work could incorporate XAI and real-time detection capability
Rallapalli, et.al. (2023)	Kaggle dataset	Bird Mating Optimization Algorithm (BMOA) for balancing, LS-SVM and LV-SVM for two-stage ensemble classification	Achieved an accuracy of 98%, effectively reducing misclassifications and associated costs compared to state-of-the-art methods.	Explore scalability and automation in real-world environments.
Xiuguo et.al. (2022)	Chinese listed companies' annual reports (MD&A textual data +	LSTM, GRU, textual feature extraction using word vectors, comparison with traditional ML approaches	Achieved classification rates: LSTM – 94.98%, GRU – 94.62%, demonstrating textual MD&A features significantly boost fraud detection accuracy.	Extend model to multilingual datasets and other industries.

	financial indices)			

3 METHODOLOGY

This study utilizes the dataset for IEEE-CIS Fraud Detection was obtained from Kaggle and includes more than 590,000 online transactions classified as either fraudulent or lawful. In the preparation stage, missing value management, data cleansing, categorical feature encoding, and Principal Component Analysis (PCA) dimensionality reduction are all included. A 70:30 train-test split is applied after random oversampling to solve the issue of class imbalance. An Artificial Neural Network (ANN) serves as the principal prediction model, with Decision Tree, LightGBM, Logistic Regression, and CNN models also being employed for comparison analysis. The memory, accuracy, precision, confusion matrices, and F1-scores are employed to assess the model's efficacy. AI-driven decision-making may become more transparent and understandable with the application of techniques like SHAP and LIME (Local Interpretable Model-agnostic Explanations). Figure 1 provides a thorough depiction of the suggested fraud detection pipeline by showing the whole methodological framework, encompassing every stage, from gathering and preparing data to training models, assessing them, and making sure they are interpretable.

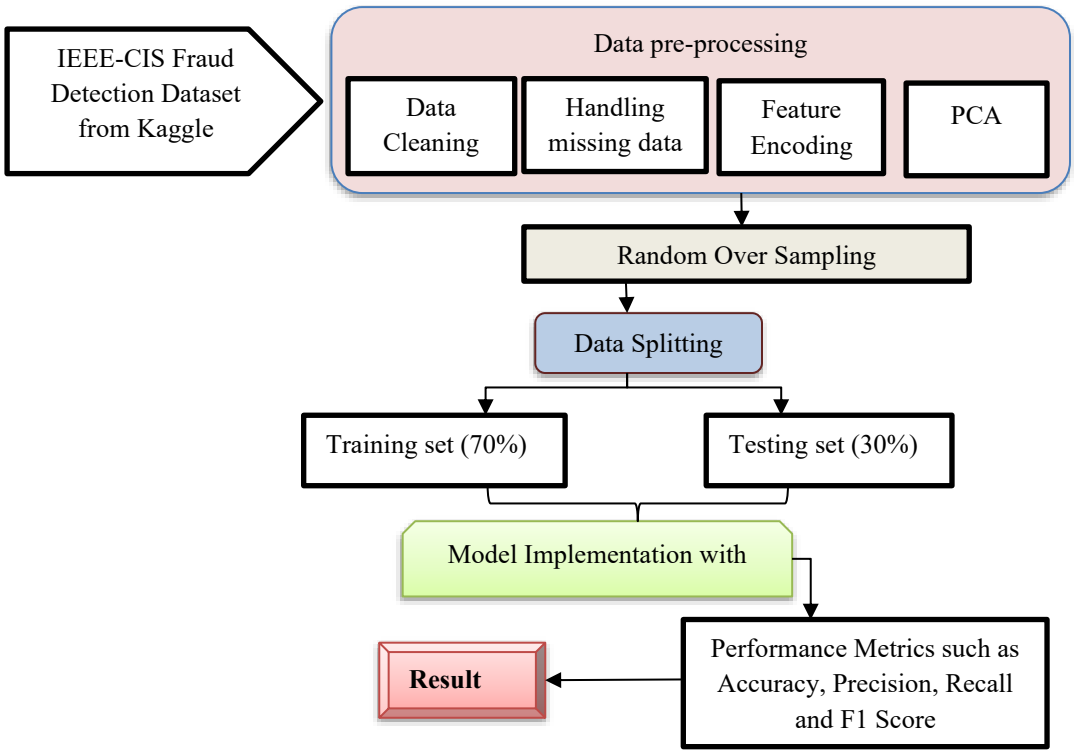

Figure 1: Proposed Flowchart for Fraud Detection

3.1 Data Collection

The IEEE Computational Intelligence Society (CIS) in collaboration with Vesta Corporation is holding a competition, a prominent payment service provider, the dataset for IEEE-CIS Fraud Detection has been made publicly accessible on Kaggle. It is among the largest and most widely used benchmark datasets for studies on fraud detection. Nearly 590,000 online transaction records, each classified as either fraudulent or legitimate, were gathered from actual e-commerce sites to create the dataset. The class distribution is illustrated in Figure 2.

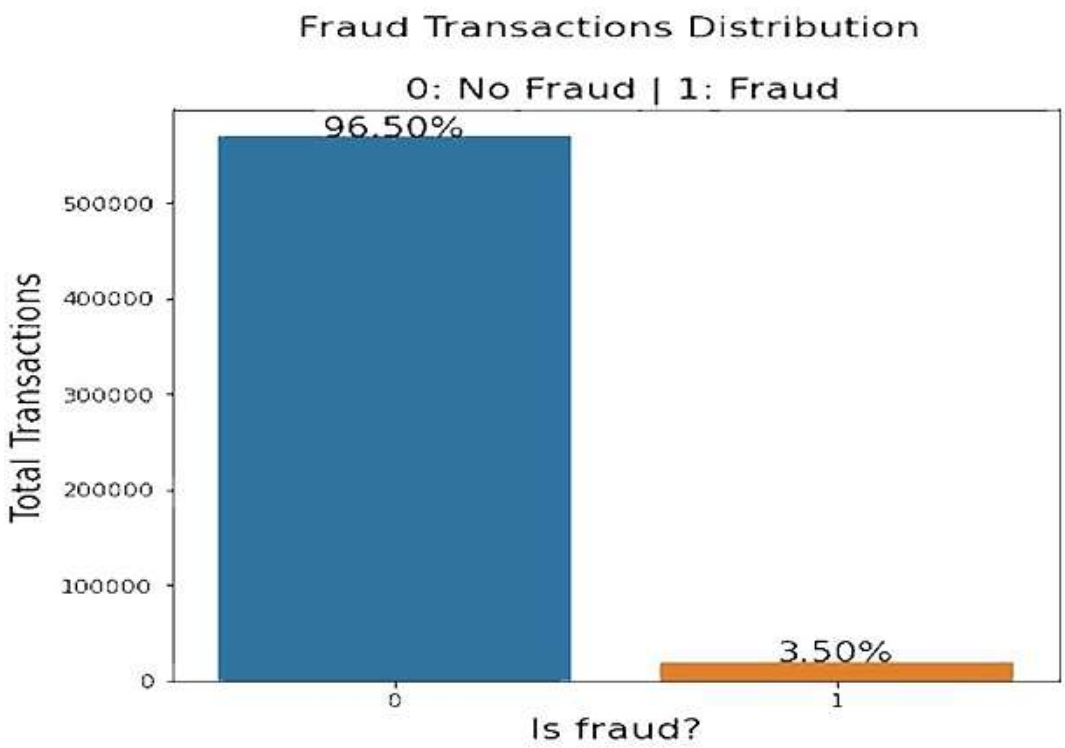

Figure 2: Distribution of Financial Transactions.

Figure 2 shows the class imbalance in a fraud dataset: the vast majority of transactions are labeled Not Fraud (class 0), comprising about 96.5% of all records, while Fraud (class 1) accounts for a much smaller 3.5%, highlighting a highly imbalanced target distribution that can affect model training and evaluation.

3.2 Dataset Preprocessing

There was preprocessing of the dataset. Data cleaning, feature encoding, and managing missing data are crucial steps. Additionally, Dimensionality reduction and handling class imbalance are necessary to ensure consistency and model compatibility. These actions were crucial for improving model performance and preventing data leaks. These steps are listed below:

- **Data Cleaning:** The practice of identifying erroneous records and corrupt data in a database table or record set is called data cleansing. The main goal of the cleaning procedure is to identify and remove inaccurate, inconsistent, irrelevant, or incomplete data, and to employ methods to alter or eliminate this useless content.
- **Handling Missing Data:** Managing a dataset's missing values is an essential part of data preparation. Inadequate data collection, system failure, or inaccurate data input are some of the causes of missing values[21]. Incomplete or corrupted entries are fixed or removed to maintain data quality and prevent bias in training.
- **Feature Encoding:** A data processing technique called categorical encoding converts categories, such colors or product kinds, into computer-manipulated integers. This facilitates improved data processing and analysis.

3.2.1 Dimensionality Reduction using Principal Component Analysis (PCA)

Data is projected onto a plane using PCA, a dimensionality-reduction approach, where each coordinate corresponds to a data feature. It then transfers this data onto a new dimension where the variation is maximized [22]. PCA is a method for creating prediction models and conducting exploratory data analysis. PCA visualizes genetic distance and is specifically made for unlabelled data. The principle A component can eliminate data noise[23].

3.2.2 Handling Class Imbalance Using Random Oversampling

The simplest oversampling technique is random oversampling. To create fresh samples in the minority class, random oversampling randomly selects samples. Even if the number of samples has risen, the produced samples are identical clones of the original samples, which might lead to overfitting because new samples are often quite close to the original samples[24]. Equation (1) is a compact equation that represents the new balanced class:

$$X_{new} = X_{minority} + X^{r}_{minority} \quad (1)$$

Where, cap X_{new} is the new dataset after oversampling, subscript is the original minority class dataset, and cap $X_{minority}$ class dataset and $X^{r}_{minority}$ represents the randomly selected rows (duplicate instances).

Figure 3: Class Imbalance using random oversampling

Figure 3 shows the Class imbalance after random oversampling, showing nearly equal frequencies (~550,000) for Class 0 (Non-Fraud) and Class 1 (Fraud), indicating successful dataset balancing.

3.3 Data Splitting

The training set and the testing set are the two halves of the dataset. It is possible to guarantee generalizability and robustness by contrasting the model's performance with the testing set and the training set's expected performance on unidentified data.

3.4 Artificial Neural Network (ANN)

An ANN can handle challenges in areas like pattern recognition and game play. ANNs' basic idea is built on neuron mimics connected in different ways[25]. Multiply the value x_1 by the weights w_1 for every input. The multiplied values should then be added up Weights are used to reflect the strength of connections between neurons, which affect a neuron's output. Even when their weights are equivalent, w_1 has greater influence than w_2 due to its greater weight. The eq. (2) and eq. (3) are shown below:

$$\sum = (x_1 * w_1) + (x_2 * w_2) + \cdots + (x_n * w_n) \tag{2}$$

The row vectors of the inputs and weights are x = [x_1, x_2, …, x_n] and w = [w_1, w_2, …, w_n], respectively, and their dot product is given as equation (3).

$$x.w = (x_1 * w_1) + (x_2 * w_2) + \cdots + (x_n * w_n) \tag{3}$$

Hence, equation (2) is equal to equation (3). Also, equation (4) is mention below:

$$\sum = x.w \tag{4}$$

The outcome of applying bias b to term "z" refer to the multiplied integers. To get the appropriate output values, the overall activation function must be biased, also known as offset, equation (5) is shown in below:

$$z = x.w + b \tag{5}$$

Transform z using an activation function that is non-linear and depends on the given value. A neuron's output would be linear without activation functions, but they are necessary to introduce non-linearity. The rate of learning by the neural network is also heavily dependent on these characteristics. Equation (6) uses the sigmoid function, also called the logistic function, as its activation function, even though the perceptron's activation function is normally a binary step function.

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}} \tag{6}$$

Equation (7) is the projected value following forward propagation serves as a representation of the sigmoid activation function.

3.5 Model Evaluation

Different performance measures were used to evaluate whether it is possible to anticipate fraudulent transactions using the ACC, PRE, REC, specificity, and F1 of the ML models. The four columns of the confusion matrix, TP, TN, FP, and FN, indicate the model's performance and stand for TP, TN, FP, and FN, respectively. This reduces the quantity of fraud cases that are ignored and assesses how well the models identify fraudulent transactions and get rid of false alarms. The calculating equations for the performance measure are as follows [26]:

- **True Positive (TP):** The number of expected True Positives (TP) in the positive data set.
- **True Negative (TN):** The number of unfavorable results that were genuinely expected to be TN.
- **False Positive (FP):** A False Positive (FP) occurs when there is a high degree of predictability regarding the number of data points that fall into both the negative and positive categories.
- **False Negative (FN):** It is called a FN when the predicted number of data points is negative but the actual number is positive.

3.5.1 Accuracy

The percentage of correctly recognized samples to all samples is accuracy, which is the most sensible performance statistic. When the target classes are evenly distributed, the process is more effective. Equation (7) defines accuracy:

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \tag{7}$$

3.5.2 Precision

Precision may be thought of as a gauge of how accurate a classifier is, the ratio of all positive test findings to accurately recognized positive samples. Equation (8) defines precision:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

3.5.3 Recall

The TPR is also known as recall or sensitivity. This metric measures how well a real-world class's predictions fared in relation to the total amount of observations. It evaluates the model's ability to predict positive cases. Equation (9) is utilized to define REC:

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

3.5.4 F1 Score

The ACC and REC weighted average are the F1. Therefore, it is a number that takes into consideration both the positive and negative outcomes. When dealing with unbalanced classes, this metric is more accurate. Equation (10) is used to define the F1 score:

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{10}$$

4 RESULTS, DISCUSSION & COMPARATIVE ANALYSIS

The experiments in an Intel Core i5-1135G7 CPU (2.40 GHz) running Windows 11 were used in this study's Dell Inspiron laptop. Additionally, the implementation and testing were executed on Google Colab using Python notebooks. Table 2 displays an ANN model's effectiveness in identifying financial fraud on the IEEE-CIS dataset. The model attained an ACC of 97.56%, a PRE of 98.82%, a REC of 98.23%, and an F1of 98.52%. The model does quite well in terms of overall accuracy, but it might be much better at detecting fraudulent transactions, as shown by the limited REC and F1.

Table 2: Results of Financial Fraud Detection on IEEE-CIS Dataset

Measures	ANN
Accuracy	97.56
Precision	98.82
Recall	98.23
F1 Score	98.52

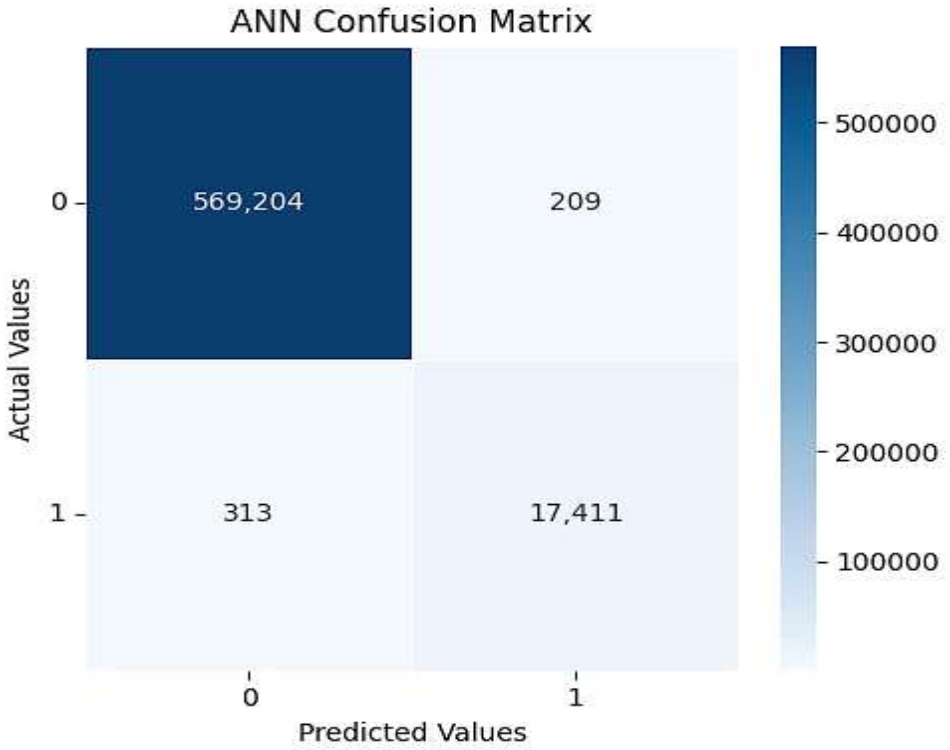

Figure 4: Confusion Matrix of ANN Model

Figure 4 displays the ANN confusion matrix, demonstrating strong performance in detecting non-fraudulent transactions with very few false positives (209). For fraudulent transactions, the model correctly identifies 17,411 cases but misses 313, indicating that while overall detection is high, minor misclassifications still occur, highlighting the importance of further improving sensitivity to fraud.

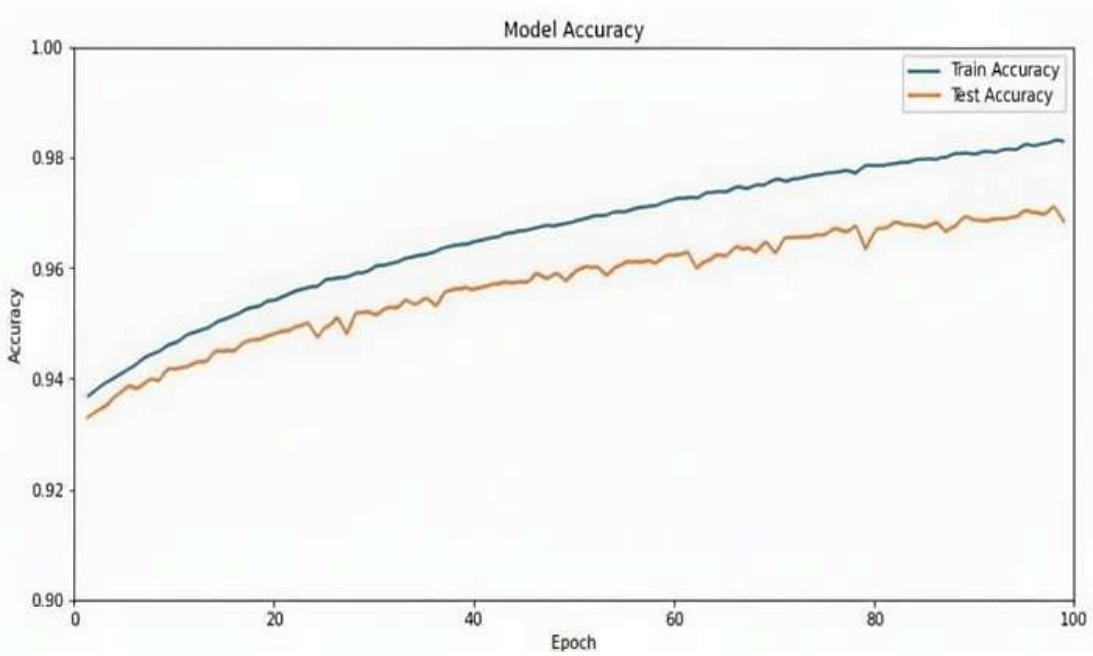

Figure 5: ANN Learning Curve of Training and Testing Accuracy

Figure 5 illustrates the accuracy curves for testing and training, which exhibit a consistent improvement across the epochs, with training accuracy reaching nearly 99% and testing accuracy stabilizing around 97%. The small gap between the two curves indicates that the ANN generalizes well, without significant overfitting, and achieves consistently high ACC on unseen data.

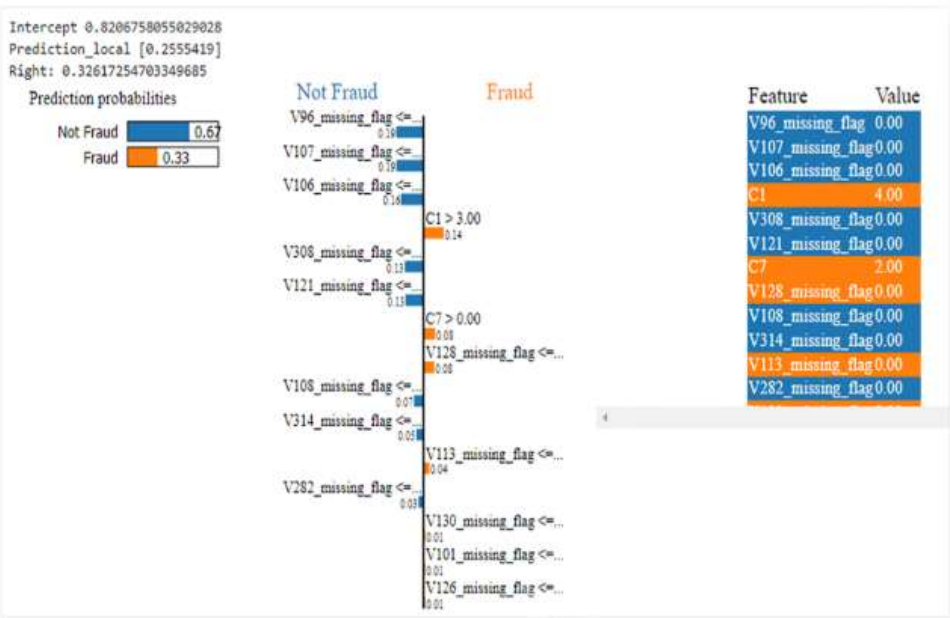

Figure 6: Lime Plot

Figure 6 illustrates the decision made by a fraud detection model. The model predicts 67% as 'Not Fraud' and 33% as 'Fraud', with features such as missing value flags favoring the 'Not Fraud' result, while higher values of C1 and C7 increase the fraud risk. Overall, the case is less likely to be fraudulent.

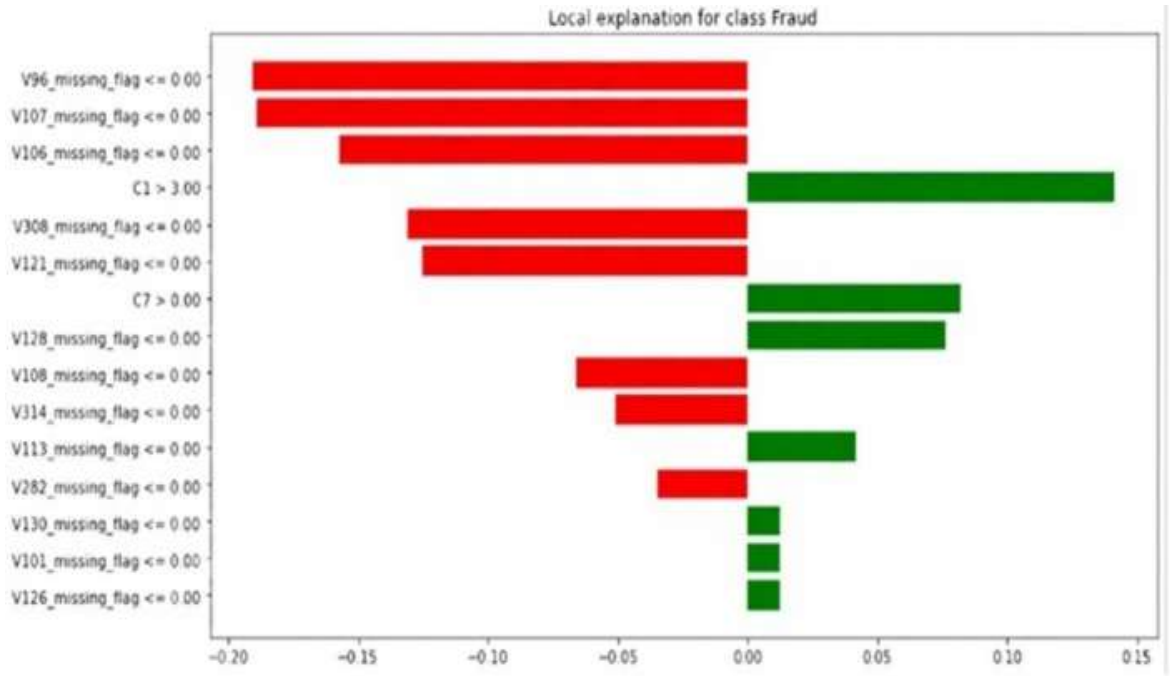

Figure 7: Lime Local Explanation Plot

Figure 7 illustrates a local explanation for predicting fraud, highlighting how individual features contribute to the decision. Features that raise the chance of fraud are shown by green bars, while those that lower it are represented by red bars. For example, missing value flags, such as V96, V107, and V106, significantly reduce the fraud probability, while higher values of C1 (> 3.00) and C7 (> 0.00) increase it. Overall, the negative contributions dominate, suggesting that the model tends to under-identify fraud, although some features still increase the risk of fraud.

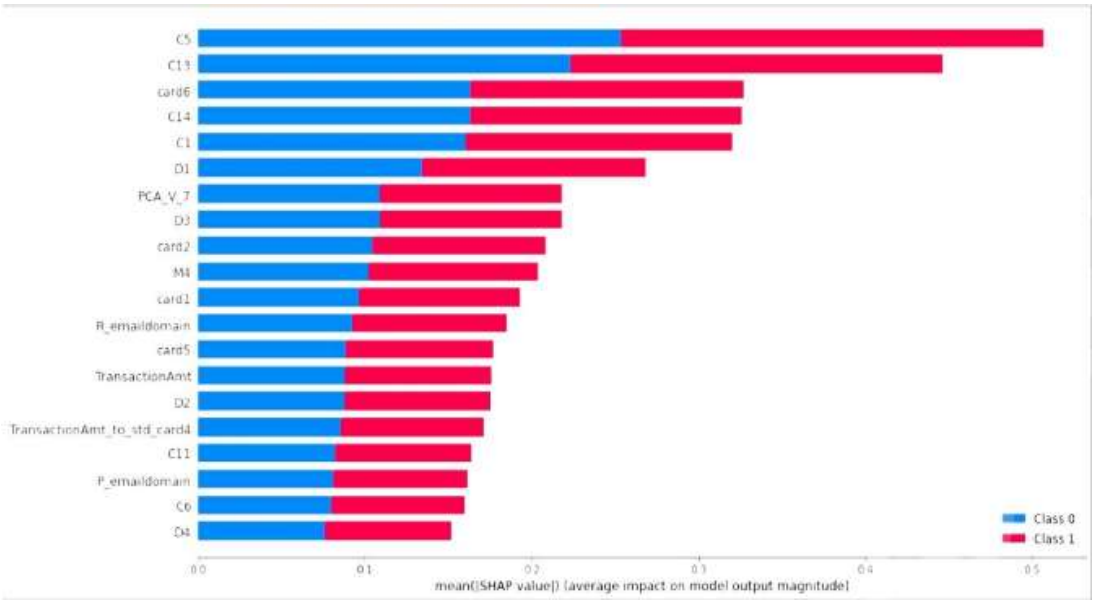

Figure 8: SHAP Summary Plot

The characteristics that most significantly affect Figure 8 displays the results of the model. Each bar represents a feature's average contribution to predicting Class 0 (blue) or Class 1 (red). Features like C5, C13, and card6 have the highest influence, making strong contributions to both classes, depending on their values. Other important features include C1, PCA_V_7, card4, and transaction-related attributes. Overall, this visualization highlights which variables most strongly drive the model's fraud classification decisions, with C5 and C13 being the most influential.

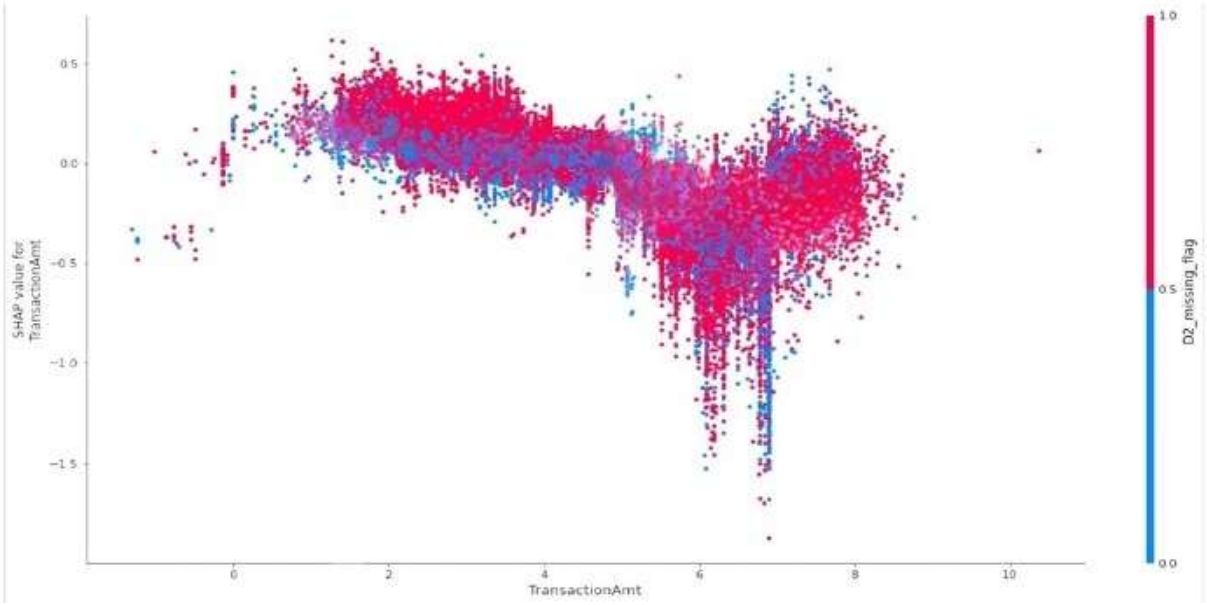

Figure 9: SHAP Dependence Plot

Figure 9 compares the Transaction Amount with a transformed feature, colored by fraud labels. Blue points are non-fraud; red points are fraud. While they overlap, fraud cases appear more concentrated in certain mid-to-high transaction ranges, indicating useful patterns for detection.

4.1 Comparative Analysis

Table 3 displays a comparison of the different models' performances. The ANN model achieves the best overall performance, demonstrating robust and balanced detection of both fraudulent and non-fraudulent transactions with 97.56% ACC, 98.82% PRE, 98.23% REC, and 98.52% F1. DT and LR show moderate performance, with accuracies of 87% and 80%, respectively, reflecting limitations in handling complex patterns. LightGBM achieves high precision (96%) but lower recall (67%), indicating it is more conservative in flagging fraud. The CNN performs well overall with balanced metrics around 92–95%.

Table 3: Comparative analysis of different models on the IEEE-CIS Fraud Detection Dataset

Models	Accuracy	Precision	Rcall	F1 Score
ANN	97.56	98.82	98.23	98.52
DT[27]	87.00	86.00	88.00	87.00
LightGBM[28]	94.00	96.00	67.00	79.00
Logistic Regression [29]	80.00	76.67	74.74	75.69
CNN[30]	95.00	92.00	93.00	92.00

The proposed ANN model outperforms all other models examined in terms of ACC, PRE, REC, and F1. Consequently, it has proven its ability to differentiate between authentic and fraudulent transactions using the IEEE-CIS dataset.

5 CONCLUSION & FUTURE WORK

The stability of the economy and public confidence in financial institutions is continuously threatened by financial fraud, making the development of accurate and transparent detection mechanisms indispensable. Traditional methods often fall short in handling evolving forms of fraud. In this study, an explainable AI framework was developed that integrates preprocessing, ANN modeling, and interpretability tools such as SHAP and LIME. Findings reveal that ANN achieves a high accuracy of 97.56, CNN delivers the most balanced results across performance metrics, and LightGBM demonstrates strong precision. The inclusion of XAI methods strengthens transparency, enabling stakeholders to understand, trust, and audit model predictions effectively. This comprehensive evaluation confirms that explainable AI has the potential to transform fraud detection into a more accountable and reliable process. Beyond performance, the integration of interpretability ensures that AI-driven fraud detection systems align with compliance standards, cultivate institutional confidence, and promote adoption in practical applications. this research contributes not only to technical advancements in fraud detection but also to building confidence in AI-enabled financial systems.
Future research directions include the development of low-latency, Systems for detecting fraud in real time that can handle big financial transactions. Enhancing recall rates for minority fraud classes remains a critical area for improvement, particularly in addressing imbalanced datasets. Expanding the framework to multi-domain and cross-border datasets would improve generalizability and scalability. Additionally, federated learning approaches can be explored for privacy-preserving fraud detection, while generative AI can simulate emerging fraud scenarios to improve robustness.

REFERENCES

[1] O. Manta, V. Vasile, and E. Rusu, “Banking Transformation Through FinTech and the Integration of Artificial Intelligence in Payments,” *FinTech*, vol. 4, no. 2, pp. 1–13, Apr. 2025, doi: 10.3390/fintech4020013.

[2] A. Parupalli, “The Evolution of Financial Decision Support Systems : From BI Dashboards to Predictive Analytics,” *KOS J. Bus. Manag.*, vol. 1, no. 1, pp. 1–8, 2023.

[3] S. Gajula, “A Review of Anomaly Identification in Finance Frauds using Machine Learning System,” *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, pp. 101–110, Jun. 2023, doi: 10.14741/ijcet/v.13.6.9.

[4] 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.

[5] A. Ali *et al.*, “Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review,” 2022. doi: 10.3390/app12199637.

[6] A. Kulal and S. Bhat, “Online Scams in Financial Market-A study with reference to Indian Financial Market,” *SSRN Electron. J.*, pp. 1–7, 2022, doi: 10.2139/ssrn.4254537.

[7] S. 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. 1–8, 2025.

[8] B. Olufemi, O. Bello, K. Olufemi, and C. Author, “Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities,” vol. 5, pp. 1505–1520, 2024, doi: 10.51594/csitrj.v5i6.1252.

[9] K. Dama, K. P. K. Reddy, K. Hrithik, D. Raheem, and Vyshnavi, “Fraud Detection in Financial Transactions,” 2024. doi: 10.13140/RG.2.2.33977.99685.

[10] S. B. Shah, “Improving Financial Fraud Detection System with Advanced Machine Learning for Predictive Analysis and Prevention,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, pp. 2451–2463, Nov. 2024, doi: 10.32628/cseit24861147.

[11] K. C. C. Hemish Prakashchandra Kapadia, “AI Chatbots for Financial Customer Service: Challenges & Solutions,” *J. Adv. Futur. Res.*, vol. 2, no. 2, p. 7, 2024.

[12] M. M. Ismail and M. A. Haq, “Enhancing Enterprise Financial Fraud Detection using Machine Learning,” *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 4, pp. 14854–14861, 2024, doi: 10.48084/etasr.7437.

[13] 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.

[14] R. Q. Majumder, “Machine Learning for Predictive Analytics: Trends and Future Directions,” *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 04, pp. 3557–3564, 2025.

[15] M. Jabeen, S. Ramzan, A. Raza, N. L. Fitriyani, M. Syafrudin, and S. W. Lee, “Enhanced Credit Card Fraud Detection

Using Deep Hybrid CLST Model," *Mathematics*, vol. 13, no. 12, p. 1950, Jun. 2025, doi: 10.3390/math13121950.

[16] S. B. Shah, "Advancing Financial Security with Scalable AI: Explainable Machine Learning Models for Transaction Fraud Detection," in *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2025, pp. 1–7. doi: 10.1109/ICDCECE65353.2025.11034838.

[17] M. Singh, N. Bansal, G. S, K.Kavithamani, M. Almusawi, and C. P. Patnaik, "Real-Time Fraud Detection in Financial Transactions Using Autoencoders," in *2024 International Conference on Advances in Computing, Communication and Materials (ICACCM)*, 2024, pp. 1–4. doi: 10.1109/ICACCM61117.2024.11058990.

[18] J. Geng and B. Zhang, "Credit Card Fraud Detection Using Adversarial Learning," in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, 2023, pp. 891–894. doi: 10.1109/ICICML60161.2023.10424872.

[19] S. Rallapalli, D. Hegde, and R. Thatikonda, "Feature Selection Based Ensemble Support Vector Machine for Financial Fraud Detection in IoT," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, 2023, pp. 1–7. doi: 10.1109/EASCT59475.2023.10392566.

[20] W. Xiuguo and D. Shengyong, "An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning," *IEEE Access*, vol. 10, pp. 22516–22532, 2022, doi: 10.1109/ACCESS.2022.3153478.

[21] S. A. Farooq, O. Konda, A. Kunwar, and N. Rajeev, "Anxiety Prediction and Analysis- A Machine Learning Based Approach," 2023, pp. 1–7. doi: 10.1109/INCET57972.2023.10170115.

[22] Z. Radeef, S. Hashem, and E. Gbashi, "New Feature Selection Using Principal Component Analysis," *J. Soft Comput. Comput. Appl.*, vol. 1, 2024, doi: 10.70403/3008-1084.1012.

[23] J. Haris Mita, C. Ganesh Babu, and M. Gowri Shankar, "Performance Analysis of Dimensionality Reduction using PCA, KPCA and LLE for ECG Signals," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1084, no. 1, p. 012005, 2021, doi: 10.1088/1757-899x/1084/1/012005.

[24] H. Ali, M. Salleh, K. Talpur, A. Ullah, A. Ahmad, and R. Naseem, "A review on data preprocessing methods for class imbalance problem," pp. 390–397, 2019, doi: 10.14419/ijet.v8i3.29508.

[25] S. Kumar, "Mathematical Model of ANN," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 3, pp. 89–93, 2021.

[26] R. Qasrawi, S. P. V. Polo, D. A. Al-Halawa, S. Hallaq, and Z. Abdeen, "Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis," *JMIR Form. Res.*, vol. 6, no. 8, pp. 1–15, 2022, doi: 10.2196/32736.

[27] F. Almalki and M. Masud, "Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods," pp. 1–30, May 2025.

[28] K. Huang, "An Optimized LightGBM Model for Fraud Detection," *J. Phys. Conf. Ser.*, vol. 1651, no. 1, 2020, doi: 10.1088/1742-6596/1651/1/012111.

[29] H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, "Credit Card Fraud Detection Based on Machine and Deep Learning," in *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, 2020. doi: 10.1109/ICICS49469.2020.239524.

[30] M. K. Pasupuleti, "Deep Learning for Fraud Detection in Real- Time Transaction Networks," *Int. J. Acad. Ind. Res. Innov.*, vol. 05, pp. 641–651, 2025, doi: 10.62311/nesx/rphcr24.

