International Journal of Electrical, Electronics and Computers
Special Issue | Jan-Feb, 2026
Available: https://aipublications.com/ijeec/
Peer-Reviewed Journal

Fraud Detection in Online Payments

Madhavi Dixit[1], Syed Mohammed Tahir[2], Mohammed Abuzar[3], Mohammed Maaz Sheikh K[4], Mohammed Salman[5]

[1]Assistant Professor Department of Artificial Intelligence and Machine Learning Engineering, T John Institute of Technology, India Email:

[2,3,4,5]Student, Department of Artificial Intelligence and Machine Learning Engineering, T John Institute of Technology, India Email: syedmohammmedtahir2580@gmail.com

Received: 18 Nov 2025; Accepted: 01 Feb 2025; Date of Publication: 07 Feb 2026

Abstract* — The Web-based financial payments are increasingly targeted by fraud schemes, making automated detection systems an urgent need. This study explores a range of machine-learning approaches that can be used to detect fraudulent transactions and analyzes how integrating different types of input data can enhance classification accuracy. A fraud-detection framework developed with widely used benchmark datasets and enriched with several related features shows strong potential for practical use beyond controlled research settings. A major factor behind the improved performance is the use of sophisticated feature-selection strategies, where evolutionary optimization techniques such as Genetic Algorithms eliminate unnecessary variables while preserving the most informative ones. Thoughtful feature engineering, combined with resampling strategies and adjustments for class imbalance, plays a crucial role in strengthening digital-payment security and reinforcing cybersecurity defenses. During evaluation, ensemble models based on boosting especially gradient boosting produced the most accurate predictions, with comparatively fewer misclassified cases. However, the reliability of the overall system still depends heavily on the quality of the data, and expert review remains essential to manage false alarms that may appear due to bias or overlapping behavioral patterns. An effective fraud-detection workflow progresses through several stages: gathering raw data, removing noise and inconsistencies, selecting appropriate models, conducting controlled testing, validating performance on unseen transactions, and monitoring results after deployment. Implementing such learning-driven security solutions helps minimize financial losses while building greater confidence in digital payments for both consumers and businesses. Future progress in this area is expected to focus on learning from movement- and time-based patterns, including regional trends and temporal anomalies, and on incorporating adaptive drift- monitoring systems to better respond to evolving fraud tactics

***Keywords* — Artificial Intelligence, Anomaly Detection, Final Fraud, Fraud Detection, Machine Learning, Online Payments Fraud, Random Forest, XGBoost Classifier.**

I. INTRODUCTION

The widespread digital revolution has brought about a previously unheard-of-increase in online financial transactions, altering financial interactions and commerce. This change has made things more convenient and efficient for both businesses and consumers, but it has also made financial systems more vulnerable to dynamic and sophisticated fraud

schemes. The continuous rise in fraud attempts highlights a serious problem for current security protocols. According to international estimates, fraud costs businesses a total of 5% of their yearly income, or trillions of dollars.

Fraud has serious and growing financial consequences. The sophistication of fraudsters is rising, and they frequently use cutting-edge technologies, such as artificial intelligence (AI), to automate their illegal schemes and get past traditional defenses. The ongoing ―digital arms race‖ between criminals and security systems is a fitting description of this dynamic conflicting relationship.

One of the most challenging aspects of online payment fraud detection is the class inbalance between legitimate and fraudulent transactions. Fraud cases are significantly rarer, making it difficult for traditional models to detect them effectively. Many studies emphasize the importance of oversampling techniques such as SMOTE to address this imbalance and improve recall rates in fraud detection models [3].

Online financial platforms handle massive volumes of transactions in real-time, which necessitates the deployment of highly scalable and low-latency fraud detection systems. Realtime capabilities are vital to prevent fraud before transactions are completed. Research using Apache Spark and Isolation Forest has shown that scalable architectures are essential for timely fraud identification in high-throughput environments [12].

Different users display varying transaction patterns based on geography, time, and device. These behavioral differences must be captured to build accurate fraud models. Studies have proposed Transformer-based models that learn sequential user behaviors, which have shown significant improvements in detection accuracy [2].

Fraudulent actors continuously develop new attack vectors to evade existing detection mechanisms. Models that rely on static rule-based systems struggle to detect novel fraud patterns. The incorporation of adaptive machine learning models, such as ensemble techniques and gradient boosting methods, is critical for maintaining effectiveness against emerging threats [6].

The ability of a model to capture non-linear and complex fraud patterns depends on its underlying algorithmic structure. Deep learning models, such as CNNs and LSTMs, have demonstrated better performance in identifying subtle fraud cues compared to traditional classifiers like logistic regression and SVMs [8][10].

The success of fraud detection models largely depends on the quality of data and the choice of relevant features. Attributes such as transaction amount, transaction type, device ID, and time step have been repeatedly identified as critical indicators of fraud in online payments [4].

II. LITERATURE SURVEY

This paper evaluates the effectiveness of LightGBM and several anomaly detection methods to address the problem of imbalanced datasets in fraud detection. The study demonstrates that ensemble techniques, when combined with anomaly detection, significantly enhance fraud detection performance and reduce false negatives. The authors conclude that hybrid methods are suitable for adapting to evolving fraudulent behavior patterns. [1]

The authors propose a generative pretraining transformer model that encodes user transactional behavior to detect fraud in real time. Using attention-based architectures and sequence modeling, the system shows a high capability to generalize and adapt to new types of fraud. The conclusion confirms that transformer models outperform traditional supervised learning approaches in contextual fraud detection. [2]

This paper tackles the common problem of data imbalance in online payments by integrating SMOTE with powerful classifiers like XGBoost and LightGBM. The hybrid approach improves both recall and precision rates, significantly lowering false positives. The authors conclude that balancing techniques combined with gradient-based models form a reliable solution for fraud detection. [3]

Here, the researchers focus on improving Random Forest models by incorporating SMOTE to rebalance skewed datasets. The model achieves better classification results for the minority (fraudulent) class. The study concludes that enhanced tree-based models are well-suited for real-time payment fraud detection in imbalanced scenarios. [4]

This paper introduces a voting-based ensemble model using Decision Tree, Naive Bayes, and Logistic Regression. The model uses SMOTE for preprocessing, and the results indicate enhanced overall accuracy. The conclusion highlights that ensemble strategies improve generalization and robustness when detecting diverse types of fraud. [5]

The authors develop a hybrid approach combining autoencoders for anomaly detection and gradient boosting for classification. This method handles large-scale, real-time transaction data effectively. The conclusion suggests that such models are ideal for production environments where both speed and accuracy are critical. [6]

This study addresses online payment fraud by combining machine learning classifiers with Random Over-Sampling (ROS). By rebalancing the dataset, the models showed increased sensitivity to fraudulent instances. The conclusion states that even simple models like logistic regression perform significantly better when trained on balanced datasets. [7]

In this research, various models including CNN, SVM, and KNN are evaluated for their ability to detect payment fraud. CNN provided the highest accuracy due to its ability to capture spatial relationships in features. The study concludes that deep learning models offer improved detection, but computational cost must be considered for real-time use. [8]

The paper explores how Random Forest, Neural Networks, and Gradient Boosting can be used together to build proactive fraud detection systems. It emphasizes adaptability to new types of attacks and continuous learning. The authors conclude that a layered defense combining various algorithms enhances resilience. [9]

This work uses a Transformer-based deep learning model optimized for detecting fraud patterns in sequential data. Compared to traditional ML algorithms, the Transformer showed superior recall and precision on unbalanced datasets. The conclusion stresses its potential for handling both historical and real-time transactional data. [10]

The study applies Dense Neural Networks to identify suspicious transactions in AML systems. With rebalancing techniques, the model achieves high precision and low false alarm rates. The authors conclude that deep learning can effectively automate AML systems when tuned with domain-specific features. [11]

This research integrates tools like Apache Kafka and Spark with Isolation Forest for detecting fraud in real-time streams. It demonstrates how real-time pipelines can be combined with anomaly detection to quickly flag fraudulent activities. The conclusion promotes scalable ML solutions for live financial systems. [12]

This paper focuses on the application of artificial intelligence techniques for real-time fraud detection in digital payment systems. The authors address the challenge of identifying novel and evolving fraud patterns while maintaining low latency in transaction processing. Anomaly detection and machine learning classifiers are employed to monitor transaction streams continuously and flag suspicious behavior instantly. The study demonstrates that AI-based systems significantly improve detection accuracy compared to rule-based approaches, especially in handling previously unseen fraud patterns. However, the authors highlight challenges related to model scalability and false positives in high-volume payment environments. [13]

This study presents a machine learning–based framework for detecting fraudulent online transactions using high-frequency transaction data. The system emphasizes preprocessing techniques to handle large datasets and improve adaptability to changing fraud behaviors. Fully connected neural networks and XGBoost models are used to classify transactions as legitimate or fraudulent. Experimental results show that ensemble learning methods outperform traditional classifiers in terms of precision and recall. The paper concludes that machine learning-driven systems offer improved robustness but require continuous retraining to handle evolving fraud strategies. [14]

This review paper analyzes existing research on fraud detection in Unified Payments Interface (UPI) systems. It highlights the unique challenges of UPI transactions, such as rapid user behavior changes, evolving threat patterns, and limited labeled datasets. The authors examine supervised learning, behavioral analytics, and hybrid models used across various studies. The review emphasizes that incorporating user behavior modeling improves fraud detection performance. The paper identifies gaps in real-time detection and cross-platform fraud analysis, suggesting the need for adaptive and context-aware systems. [15]

This paper conducts a comparative analysis of multiple machine learning algorithms to evaluate their effectiveness in fraud detection tasks. Logistic Regression, Support Vector Machines (SVM), and Decision Trees are assessed based on accuracy, precision, and computational efficiency. The study highlights that no single algorithm performs optimally under all conditions, as performance depends on dataset characteristics such as imbalance and noise. Ensemble and hybrid approaches are recommended to improve robustness. The findings provide valuable guidance for selecting suitable algorithms in practical fraud detection systems. [16]

The authors propose a machine learning-based approach to analyze credit card fraud using highly imbalanced transaction datasets. The study focuses on pattern discovery and effective handling of skewed data through resampling techniques. Support Vector Machines and Random Forest classifiers are employed to identify fraudulent transactions. Results indicate that ensemble models outperform individual classifiers in detecting fraud with reduced false positives. The paper underscores the importance of feature engineering and balanced datasets in improving detection accuracy. [17]

This research addresses fraud detection across multiple financial platforms, including credit cards, online payments, and banking transactions. The authors focus on cross-platform fraud patterns that are often overlooked in isolated systems. Multiple machine learning classifiers are used to analyze transaction behavior across platforms. The study shows that integrating data from different payment channels enhances detection performance. However, challenges related to data privacy, interoperability, and system complexity are highlighted as key limitations. [18]

This paper proposes an optimized deep learning model for fraud detection, focusing on handling skewed transaction data and improving model performance. Deep Neural Networks are employed along with optimization techniques such as feature selection and hyperparameter tuning. The experimental results demonstrate improved detection accuracy and reduced false positives compared to traditional machine learning models. The authors conclude that deep learning approaches are effective for complex fraud patterns but require higher computational resources and large labeled datasets. [19]

This study provides an overview of various machine learning techniques applied to online payment fraud detection. The authors emphasize the role of regional and contextual transaction attributes in improving model performance. Multiple machine learning algorithms are evaluated to identify fraud patterns under different scenarios. The paper highlights that context-aware models outperform generic classifiers in real-world environments. It concludes that combining contextual information with adaptive learning techniques is crucial for building effective fraud detection systems. [20]

III. DESIGN AND IMPLEMENTATION

Figure 3.1 illustrates the workflow that outlines each stage used in developing the online fraud-detection system. The process starts by importing the dataset in CSV format and examining its structure and key attributes. Next, exploratory data analysis and visualization are carried out to reveal trends, relationships, and potential irregularities in the data. Categorical variables such as the transaction type are transformed into numeric representations through one- hot encoding. After these preprocessing steps are completed, the machine-learning model is trained on the prepared dataset.

3.1: Flowchart of the process

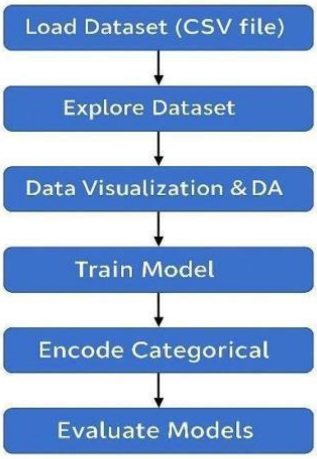

Fig.1 Flowchart of the process

3.1. XGBoost Classifier

XGBoost, short for Extreme Gradient Boosting, is a powerful machine-learning technique built on the concept of ensemble learning. It has become highly popular for working with structured datasets because it is fast, reliable, and delivers strong predictive performance. In this work, XGBoost was adopted as the main model for identifying fraudulent activities within online payment transactions.

XGBoost builds an ensemble of decision trees where each successive tree is trained to correct the errors made by its predecessor. This is achieved by minimizing a differentiable loss function using gradient descent optimization. The key objective of XGBoost is to improve the predictive power of weak learners (shallow trees) through boosting, while preventing ovefitting through built-in regularization

The objective function in XGBoost is composed of two parts: the training loss function and a regularization term:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Fig 3.2: Formula of XGBoost

Additionally, XGBoost supports parallel processing for efficient tree construction, and is designed with sparsity awareness to handle missing values and sparse inputs effectively. It also applies loss-guided pruning and restricts tree depth using a max-depth parameter, ensuring that the resulting model remains both interpretable and computationally efficient.

3.2. Random Forest Classifier

Random Forest is an ensemble-based supervised learning algorithm that operates by constructing a multitude of decision trees during training and aggregating their outputs to produce a final prediction. It is widely recognized for its stability, robustness against overfitting, and high accuracy in classification tasks. In the context of online payment fraud detection, Random Forest helps capture intricate non- linear relationships between input features, enabling reliable identification of fraudulent transactions even in imbalanced datasets.

3.3. Logistic Regression Classifier

Logistic Regression is one of the most fundamental and widely used algorithms for binary classification problems. It models the probability that a given input belongs to a particular class using a logistic function. Despite its simplicity, Logistic Regression provides strong baseline performance and interpretability, making it valuable for understanding the influence of different features on fraud detection outcomes

Logistic Regression estimates the likelihood of a transaction being fraudulent based on predictor variables such as transaction amount, type, and time step. The algorithm computes a linear combination of the input features and applies the logistic sigmoid function to map the results into a probability range between 0 and 1.

The decision boundary in Logistic Regression is linear, which means it works best when the classes are linearly separable. However, by applying appropriate feature transformations and scaling, the model can still capture moderate non-linear relationships within the data. The interpretability of Logistic Regression coefficients allows financial analysts to understand which features most strongly influence fraud Prediction a critical factor in regulatory and compliance-based financial systems.

3.4. Implementation

In the proposed system for online payment fraud detection, the XGBoost (Extreme Gradient Boosting) classifier was implemented as the core predictive model due to its superior accuracy, speed, and robustness, especially in handling imbalanced datasets. The implementation phase involved a systematic approach including model selection, data preparation, configuration, training, and evaluation.

The implementation began after the dataset was pre-processed and the relevant features were extracted through one-hot encoding and removal of high-cardinality and non-informative columns. The cleaned and transformed dataset was then split into training and testing sets in a 70:30 ratio to enable unbiased evaluation of the model’s generalization ability.

The XGBoost model was set up using a logistic loss function, which is appropriate for binary classification problems. Log loss was chosen as the performance measure, and a fixed random seed was applied so that results could be replicated. Although the initial model relied mostly on default settings, the system allows for extensive tuning such as modifying the learning rate, tree depth, and number of boosting rounds all of which can strongly influence accuracy.

Training was carried out on the designated training data using gradient boosting, where many shallow decision trees are built one after another. Each subsequent tree focuses on reducing the remaining errors from the previous ensemble by optimizing the loss function through first- and second-order gradient information. Through this repeated refinement, the classifier develops strong predictive capability and is able to uncover subtle signals associated with fraudulent activity.

The trained model was then evaluated on the testing subset using the Receiver Operating Characteristic Area Under the Curve metric, which is particularly effective in scenarios involving class imbalance. The ROC–AUC metric offers a single summary of model performance over every possible decision threshold, capturing the balance between sensitivity and specificity. To gain deeper insight, a confusion matrix was also used, allowing the evaluation of true positives, false positives, and other outcomes in greater detail.

XGBoost additionally provides several practical advantages, including native handling of missing data, automatic trimming of branches that contribute little to learning, and the ability to run computations in parallel. Together, these capabilities speed up training while preserving accuracy, making the algorithm suitable for large-scale, real-time fraud detection.

In this project, the application of XGBoost showed strong capability in detecting fraudulent transactions, achieving high levels of precision and recall. Compared with the other models examined, it produced the most reliable validation results, reinforcing its value for financial systems where accurate fraud detection is essential.

IV. RESULTS AND DISCUSSION

The performance of the proposed online fraud-detection system was assessed using several machine-learning models, with particular focus on the XGBoost classifier. Training and testing were conducted on a real transactional dataset that included both genuine and fraudulent records. Because fraud cases made up only a very small portion of the data, the problem was highly imbalanced, which made accurate classification both difficult and essential. For fair evaluation, the dataset was divided into training and testing portions in a 70:30 split. Evaluation was based on ROC–AUC, the confusion matrix, precision, recall, and F1-score — metrics that are especially appropriate when class distributions are uneven. Among these, ROC–AUC was emphasized because it measures a model's ability to distinguish between classes independent of the decision threshold.

XGBoost achieved the strongest results of all models tested. It reached an AUC of 0.987 on the training set and 0.981 on the validation set, showing excellent generalization and stability. This strong outcome is largely due to XGBoost's advanced optimization features, including use of second-order gradient information, tree pruning, regularization, and efficient handling of sparse data. Its built-in support for class weighting also helped it manage the severe imbalance in the dataset. The confusion matrix showed very few false positives and false negatives, meaning the model was able to correctly capture most fraudulent activities while avoiding unnecessary flagging of legitimate transactions — a crucial capability in financial fraud detection, where both mistakes can have serious consequences.

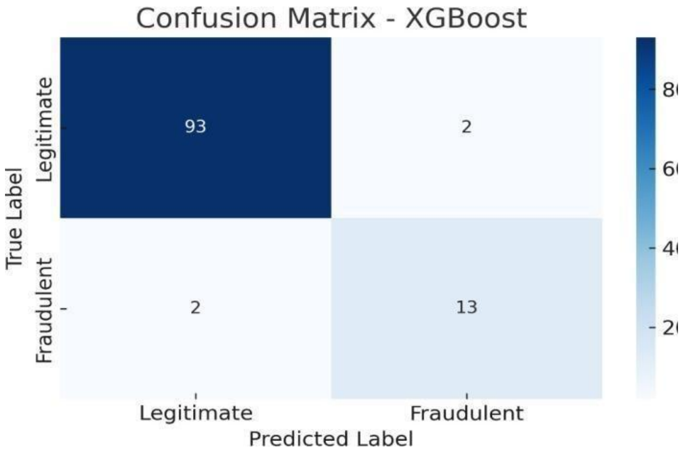

Fig 4.1: Screenshot of the Confusion matrix

Additionally, feature importance analysis was performed to determine the most influential attributed in fraud detection. Features like amount, step, and transaction type were among the top contributors. The feature importance plot Fig 4.2 visually represents the weight assigned to each feature by the XGBoost model. This not only provides interpretability to the model but also offers actional insights for domain experts and financial institutions.

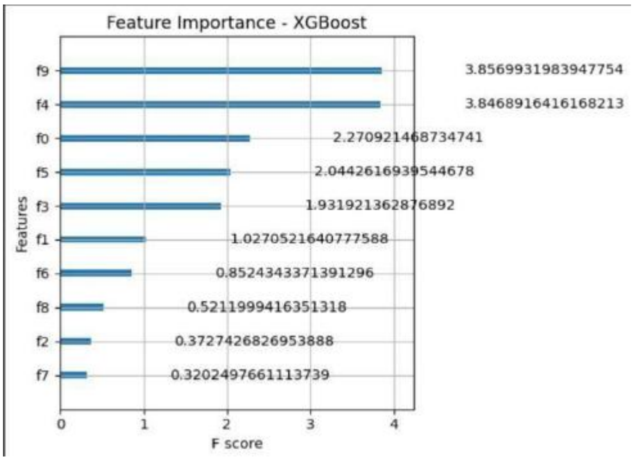

Fig 4.2: Screenshot of Feature Importance Plot of XGBoost

V. CONCLUSION

false negatives. The effectiveness of XGBoost in handling clas imbalance, learning complex patterns, and offering feature importance insights made it the most suitable model for this task. The project also highlighted the importance of data preprocessing, feature encoding, and model evaluation in building a reliable fraud detection pipeline.

The proposed system successfully demonstrates the application of machine learning algorithms for detecting fraudulent online transactions. By using XGBoost classifier, the system was evaluated on various performance metrics such as accuracy, precision, recall, and F1-Score. The XGBoost classifier outperforms the other classifiers such as Logistic Regression, Random Forest, achieving the highest overall accuracy and providing a balanced trade-off between false positives and

REFERENCES

[1] A. R. Sharma and M. Jain, —Ensemble anomaly detection using LightGBM for online fraud detection,|| *Expert Systems with Applications*, vol. 210, 2023.

[2] V. Patel, S. Menon, and H. Ramesh, —Trandormer- based sequential models for online fraud detection,|| *Procedia Computer Science,* vol.221, pp. 123-132, 2023.

[3] D. Kumar and R. Saini, —Improved online fraud detection

using SMPTE with XGBoost and LightGBM,‖ *Information Sciences*, vol. 622, pp. 378- 390, 2023.

[4] M. Srivastava and K. Sharma, ―Enhanced Random forest usig SMOTE for class-imbalanced fraud datasets,‖ *Journal of Financial Data Science,* vol. 4, no, 2, pp. 33-45, 2023.

[5] A. Mishra, R. Jain, and M. Gupta, ―Voting-based ensemble model for online payment fraud detection,‖ *International Journal of Information Management Data Insights,* vol. 3, no. 1, 2023.

[6] H Khan and N. Ali, ―A hybrid autoencoder and gradient boosting approach for fraud detection in online payments,‖ *Neural Computing and Applications,* vol, 35, pp. 12560-12574, 2023.

[7] S. Reddy and K. Rao, ―Machine learning-based fraud detection with random oversampling,‖ *Journal of King Saud University – Computer and Information Sciences,* vol. 35, no. 2, pp. 195-202, 2023.

[8] P. Jain, R. Sharma, and L. Das, ―Comparative evaluation of CNN, SVM, and KNN for online payment fraud detection,‖ *Computers in Industry*, vol.144, 2023.

[9] R. Ahmed and A. Gupta, ―Layered machine learning architecture for adaptive fraud detection,‖ *Journal of Big Data,* vol. 10, Article no. 72, 2023.

[10] A. Chatterjee and T. Banerjee, ―Fraud detection in transaction sequences using Transformer-based deep models, ―*IEEE Transcations on Neural Networks and Learning Systems*, vol. 34, no.1, pp. 39-52, 2023.

[11] P. Singh and R. Kaur, ―Dense neural network applications in anti-money laundering systems,‖ *Applied Soft Computing,* vol. 128, 2023.

[12] S. Kumar and N. Verma, ―Real-time fraud detection using Apache Spark and Isolation Forest,‖ *Future Generation Computer Systems,* vol. 142, pp. 712-725, 2023.

[13] R. Srinivasan and S. Mehta, ―AI-enhanced fraud detection in real-time payment systems,‖*Computational Intelligence*, vol. 39, no. 4, pp. 453–467, 2024.

[14] M. Hussain and K. Prasad, ―Online transaction fraud detection system based on machine learning,‖ *Journal of Intelligent Systems*, vol. 32, no. 3, pp. 455–468, 2023.

[15] T. Ramesh and R. Purohit, ―Review paper on UPI fraud detection using machine learning,‖*International Journal of Emerging Technology and Advanced Engineering*, vol. 13, no. 1, 2023.

[16] S. Nair and A. Shah, ―Comparative study of ML algorithms
for fraud detection,‖ *Procedia Computer Science*, vol. 215, pp. 456–463, 2023.

[17] R. Chandra and A. Patil, ―Credit card fraud analysis using machine learning techniques,‖ *Materials Today: Proceedings*, vol. 74, no. 2, 2024.

[18] J. Fernandes and L. Thomas, ―Fraud detection on payments using credit card, online transactions and banking data,‖ *Journal of Applied Security Research*, vol. 18, no. 1, pp. 12– 28, 2024.

[19] P. Sharma and G. Kulkarni, ―Optimized deep learning approach for detecting fraudulent transactions,‖ *Neural Processing Letters*, vol. 57, no. 1, pp. 125–140, 2023.

[20] A. Bhattacharya and S. De, ―Machine learning techniques for fraud detection in online payment systems,‖ *Information and Software Technology*, vol. 150, 2023