**Hybrid Machine Learning for Context-Aware Personalized Fraud Detection : Pre and Post Transactions Analysis**

Hazeera Babu a,\*

a,\*Department of Computer Science and Engineering , Chennai Institute of Technology, Kundrathur , Chennai,

Tamilnadu, India- 6000069

(Corresponding author email – hazeerab.cse2023@citchennai.net)

***Abstract*—** **Digital financial transactions have become incredibly rapid, making fraud detection highly advanced and needed in protection of consumers and the financial institutions. This paper will introduce a hybrid model in machine learning that focuses on a context-aware and personalized model in fraud detection by using anomaly detection pre-transactional activities and post-transaction classification. The proposed system utilizes user-specific behavior baselines, Isolation Forest for anomaly detection, PCA for the reduction in dimensionality, and also multi-task neural network on fraud classification. Performance of a model is evaluated which demonstrates a system's ability with regard to efficacy, giving an accuracy of 94.67% for the task pre-fraud and 94.7% for a post-fraud task that both ROC-AUC results exceed 95%It's a fraud-detection model which can help address problems such as data imbalance and dynamic patterns of fraud. This work is strong and adaptable, providing an effective solution to modern financial ecosystems, mainly in security, accuracy, and reliability. Some additional improvements like explainable AI and real-time adaptability are further discussed to polish this innovative approach.**

Keywords— Financial Fraud Detection, Machine Learning, Neural Networks, Anomaly Detection, Multi-Task Learning

# Introduction

*A. Background:*

Financial fraud is becoming an evolving global issue where fraudsters can target digital payment systems and Internet transactions. The volume of digital financial transactions has attracted many cybercrime opportunities that capitalize on various vulnerabilities [1]. From the evolution of financial transaction from online, fraud moves have developed, thus becoming more significant in cases where data breach, unauthorized transfer, and identity theft come into play. These trends are huge threats to both financial institutions and their customers, making effective fraud detection solutions a matter of urgency These trends pose significant threats to both financial institutions and their customers, underlining the urgent need for effective fraud detection solutions [2].

*B. Literature Review*

Financial fraud detection has gained much focus in research in recent times because of the critical nature of this theme in making sure that digital financial transactions have security. Over the last few years, several techniques emerged to determine fraudulent activities based on the rule-based systems, through machine learning to deep learning models. Many challenges exist in this theme, starting from adapting a new pattern of fraud with model interpretability.

*Existing Fraud Detection Methods*

1.Rule-Based Systems:

The first fraud detection systems were mainly rule-based and used manual definitions of rules and heuristics. These systems flagged suspicious transactions based on predefined criteria such as the amount of transaction, frequency, and location. Even though they were effective at the initial stages of fraud detection, these systems had several limitations. First, they were not flexible and hence could not adapt to changing fraud patterns as fraud tactics evolved over time. Second, they had high false positives because the predefined thresholds to detect fraud often flagged legitimate transactions as fraudulent. Lastly, rule-based systems were narrow in scope; they could not learn anything from historical data or identify new fraud schemes. Hence, they performed less efficiently as fraudsters kept on changing their strategies[3].

2. Traditional Machine Learning (ML) Models

In an attempt to correct the deficiencies of rule-based systems, ML models such as decision trees, random forests, and support vector machines (SVM) have been used to detect fraud[4]. These models provide a higher level of flexibility through the automatically learned patterns of data so that they adjust better according to changes in fraudulent behaviors. Nevertheless, there are still other problems in its implementation. For one, this model highly relies on the availability of sufficient data and large data samples to make the required annotations in fraud detection. They are also hindered by an imbalanced data set that poses the rare transaction as being a fraction of genuine transactions. This results in skewed prediction where it will most of the times over-estimate the correct transaction but is short of estimates fraudulent one[5].

3. Deep Learning Models:

The advances of deep learning bring even sophisticated techniques like neural networks. This technique is seen as to promise significant advancements in detection with the presence of intricate anomaly patterns and trends of data being highly associated with fraud transaction datasets[6].Common models include the usage of autoencoders, convolutional neural networks, CNNs, and recurrent neural networks, RNNs. Such models are quite apt for processing huge and high-dimensional data and, hence, more capable of managing the intricacies of financial transactions[7]. Deep learning models come with various problems. A major issue here is the computational needs. These models need significant amounts of resources and massive datasets to be trained, which can often create a hindrance in its deployment in real-time environments where resources may not be sufficiently rich. Moreover, many of the deep learning models are called "black-box" models wherein it's hard to decipher or to explain their decisions. As a result of this problem, understanding how a transaction is marked fraudulent is indispensable for developing a sense of trust in a system.

Table 1 describes which include rule-based systems, traditional machine learning models, and deep learning models. The advantages and challenges of each method are pointed out to show the progress from static rule-based approaches to more flexible but complex deep learning techniques.

Table 1. Comparison of Recent Fraud Detection Methods

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Description** | **Advantages** | **Challenges** |
| **Rule-Based Systems** | Predefined rules and thresholds to flag suspicious transactions. | easy to implement | adapt to new fraud patterns; high false positives. |
| **Traditional ML Models** | models like Decision Trees, Random Forests, and SVM to learn patterns from data. | adapts to changing patterns | Trouble with imbalanced data; requires a large annotated dataset. |
| **Deep Learning Models** | Neural networks (Autoencoders, CNNs, RNNs) for complex anomaly detection. | Manage big, high-dimensional data | High cost; lack of interpretability |

*C. Motivation:*

The need for real-time and post-transaction fraud detection has never been more intense. Traditional fraud detection systems face difficulties in keeping up with the changing nature of fraudulent activities and therefore result in delays in detecting and preventing fraud. Real-time detection is necessary to prevent fraud at the point of transactions, thereby minimizing losses and maintaining trust in digital payment systems. However, the post-transaction fraud detection is equally important because it will help in identifying and analyzing fraud that may not be visible at first glance. A good fraud detection system must integrate both approaches to provide comprehensive security.

*D. Objective:*

This research explores a new concept in detection of financial fraud by personalizing and employing multi-task learning. The developed model takes user-specific baselines, where it monitors the individual behavior of every transaction dynamically through the usage of features like average values of transactions, frequency of transactions, and patterns established. It brings together a pre-transaction analysis process for real-time fraud detection and post-transaction analysis to evaluate its outcomes comprehensively. The model leverages multi-task learning to generate predictions of pre-fraud and post-fraud scenarios. Therefore, these patterns across the tasks assist in the model's capabilities to detect improvements in their performance. Overall, there are contributions of a secure, efficient, and reliable fraud detection system that supports the reduction of risks when financial transactions occur in a real-time setting or else after transactions.

*E. Limitations of Current Techniques:*

Though all the approaches mentioned above significantly contributed toward the area of fraud detection, yet numerous limitations have been present to limit these techniques. The key problem is adaptability to changing fraud patterns. Most rule-based systems and traditional models of machine learning fail to evolve with new fraud tactics. Since fraudsters are changing their strategy continuously to circumvent detection systems, a static model loses its efficacy over time. Class imbalance is the other major challenge where the ratio of fraud transactions to genuine ones is highly biased. This imbalance creates a bias toward predicting legitimate transactions, which can result in missed detection of fraudulent activities. Lastly, interpretability issues remain a concern with deep learning models. While these models are powerful, their "black-box" nature makes it difficult to explain the rationale behind fraud detection decisions. This lack of transparency is problematic, especially in regulated environments where understanding the decision-making process is crucial for regulatory compliance and building trust among stakeholders.

How the Proposed Model Overcomes These Limitations:

1. Personalized Fraud Detection: The use of user-specific baselines renders the system more adaptable and accurate. Since the model learns to understand the transaction behavior of individual users, it becomes more effective at identifying the fraudulent activities related to those specific patterns, which does not happen in a rule-based system.

2. Handling data imbalance: The model uses advanced techniques such as SMOTE to augment the data and generate synthetic samples of fraudulent transactions, hence improving the model's ability to detect fraud despite the imbalance in the data.

3. Mixing both Pre- and Post-Fraud Analysis: The dual model of fraud detection in a real-time (pre-fraud) and post-transaction (after the transaction) increases its accuracy in detecting fraud because it uses past fraud patterns to make predictions, while post-fraud analysis increases its robustness against fraudulent strategies.

4. Multi-Task Learning: Simultaneously solving multiple tasks, for instance, fraud detection before and after transactions, utilizes shared knowledge between tasks to improve the overall detection performance of the model. It also allows the model to learn from both pre-fraud and post-fraud scenarios, thereby giving a much more complete fraud detection system.

5. Model Interpretability: By using a combination of neural networks and traditional anomaly detection methods within the proposed model, besides enhancing performance, such a fusion helps in attaining clearer results that may be less difficult to understand, bringing the process further towards explaining the detection by stakeholders involved.

*II. Methodology*

*A. Dataset Description*

A dataset with fictive objects synthetically developed based on assistance the Faker library during this work; it aids simulating different sorts of data-which then has become so appealing for simulating testing environments against fraud detection systems due to transaction's anonymity in different cases as they are quite diverse. Transaction-related records or records that do nearly resemble one being closely obtained can further mimic real data which can hence enrich the whole designing and implementation process towards development and deployment of different forms of detection schemes.

**Table 2.Features in dataset**

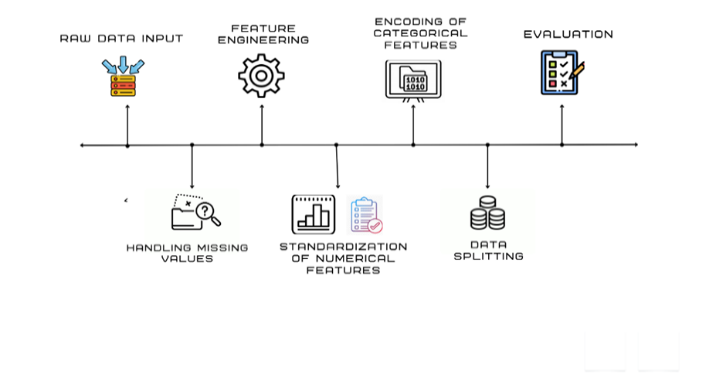
|  |  |
| --- | --- |
| **Features** | **Descriptions** |
| **User \_ID** | Unique identifier for every user. |
| **Transaction\_ ID** | Unique identifier for every transaction. |
| **Date\_ Time** | Timestamp of the transaction |
| **IP\_ Address** | The IP address from where the transaction was initiated |
| **Latitude & Longitude** | Geolocation of the transaction. |
| **Amount** | The amount involved in the transaction. |
| **Merchant** | The type of merchant or category engaged in the transaction |
| **Payment\_ Method** | The mode of payment used (credit card, debit card, UPI) |
| **Device** | From where the transaction was initiated-desktop ,mobile. Location: Where the transaction is taking place |
| **Risk\_ Profile** | Categorized level of the risk of transaction. Low, Medium, or High |
| **Suspicious** | Target variable showing if the transaction is suspicious. It is set as 1 for the suspicious one and 0 for others |
| **Transaction\_ Frequency** | How often a user performed a transaction. |
| **Location\_ Match** | A flag that matches whether the location of the transaction falls under the user's location or not. |

The dataset also has two types of features: pre-fraud and post-fraud. It will make the fraud detection process even more rich:

Pre-fraud features are basic transactional data, like Amount, Payment\_ Method, Merchant, and derived metrics like Amount\_ Deviation, which tells how unusual the amount is with respect to user historical data. ,Post-fraud features are features that are added after fraud detection has been triggered. They include metrics such as Location \_Match and other behavioral features that reflect the detection of suspicious activities.The synthetic nature of the dataset, which is created with Faker, ensures that the transactions it contains are diverse, with both legitimate and fraudulent patterns. This provides a good environment for the robust testing of fraud detection models while maintaining control over a reproducible environment. With the use of Faker, it is possible to generate large datasets, which are necessary in the training and evaluation of machine learning models for fraud detection.

*B.Data Preprocessing*

Data preprocessing is basically very important to ensure that a dataset is clean, properly structured, and thus appropriately ready for use in machine learning models. First, it addresses missing values in the dataset by either imputing or removing them, with a decision based on their importance and distribution. Feature engineering involves coming up with new variables and selecting essential anomaly patterns within a group of transactions; examples of new variable that came about in the selection included Amount \_Deviation, Frequency \_Deviation, and Hour\_ Deviation. Standard Scaler normalizes numeric features such as Amount, Latitude, and Longitude into one scale for aiding proper improvement since those ranges significantly outweigh in dominating a learner. The categorical features, such as Merchant, Payment\_ Method, Device, and Risk\_ Profile are encoded using One Hot Encoder to transform them into a form of binary variables to be fed into the machine learning models. Finally, the data sets split into training and testing sets, preferably split in a 70 -30 ratio, where the model undergoes training on a vast extent of data and is then exposed to an unseen set that determines whether the model will successfully generalise to the set.

 Figure 1. Preprocessing Pipeline

*C. Hybrid Model Architecture*

The proposed hybrid model integrates anomaly detection for pre-fraud analysis and fraud classification for post-fraud analysis within a multi-task learning framework, enhancing the accuracy and robustness of the fraud detection system.

**Pre-Fraud Module:**

The pre-fraud analysis is handled by the **Isolation Forest** algorithm, an anomaly detection technique designed to identify outliers by isolating data points in high-dimensional spaces [5]. This method is particularly effective in fraud detection scenarios, where fraudulent transactions often deviate significantly from a user's historical behavior. By applying Isolation Forest, transactions that show anomalous patterns are flagged for further investigation, facilitating early detection of potential fraud.

**Post-Fraud Module:**

Once potential fraudulent transactions are detected, the model employs a **multi-task neural network** for fraud classification[8-9]. This architecture comprises shared layers that learn common features from both pre-transaction and post-transaction data. These shared layers are followed by separate output layers for pre-fraud and post-fraud predictions, enabling the model to leverage both types of information to refine its fraud detection capabilities.

* **Shared Layers:** These layers capture common features that are relevant to both pre- and post-fraud tasks, allowing the model to learn shared patterns that improve generalization.
* **Pre-Fraud Output:** This output layer classifies transactions as suspicious or not based on the pre-fraud features, such as **Amount** and **Transaction Frequency**.
* **Post-Fraud Output:** This layer further refines the fraud prediction by incorporating post-fraud features, such as **Location\_Match**, to improve detection accuracy by considering transactional context after fraud has been flagged.

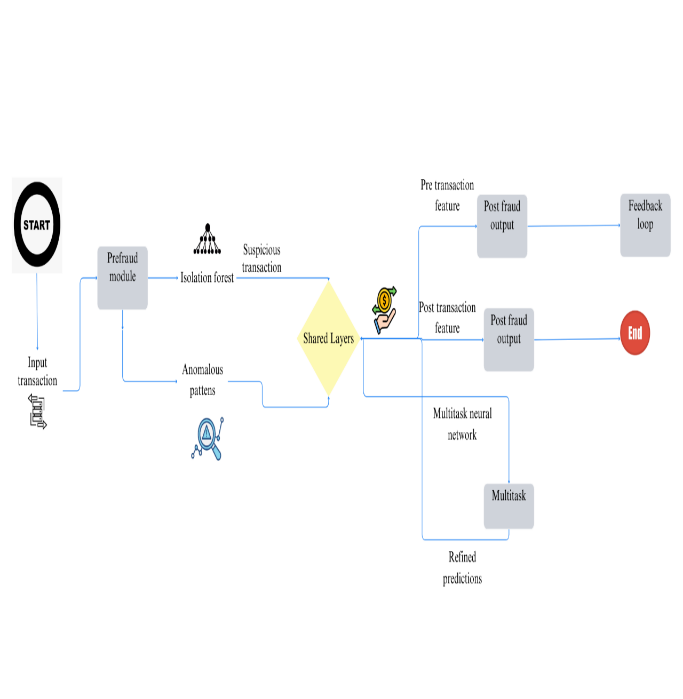


Figure 2. Hybrid Model Architecture

Figure 2 describes combining these two modules, the hybrid model provides a comprehensive and effective approach to fraud detection, ensuring that both immediate anomalies and broader fraud patterns are addressed.

*D. Implementation Details*

The proposed hybrid fraud detection model combines anomaly detection and fraud classification using a multi-task neural network. The model is implemented using Python as the primary programming language, and Pandas is used for data manipulation and preprocessing. For anomaly detection, Scikit-learn provides the Isolation Forest algorithm, which is used to identify outliers in the dataset. This library also features scaling tools and classification evaluation functions to measure the performance of the model. TensorFlow is used to implement and train the multi-task neural network that simultaneously handles both pre-fraud and post-fraud tasks.

**The following are some key parameters set in the training setup of the model:**

For Isolation Forest, contamination is set to 0.1, which helps identify anomalies within the dataset. The input layer is set to be equal to the number of features after preprocessing. The hidden layers of the network consist of a shared layer with 128 units and ReLU activation function followed by dropout to prevent overfitting. There are two different output layers for pre-fraud and post-fraud prediction with one unit each using sigmoid activation function for binary classification. The model uses the Adam optimizer with a binary cross-entropy loss function for both tasks, ensuring efficient learning. Training is done for 10 epochs with a batch size of 32.For the evaluation of the model, several metrics are used, including accuracy, precision, recall, F1-score, and ROC-AUC, for both pre-fraud and post-fraud classification tasks. These metrics help in assessing the model's ability to detect fraudulent transactions effectively at different stages.

**Hybrid Model Architecture**

Hybrid model architecture is used that integrates pre-fraud anomaly detection and post-fraud classification in a multi-task learning framework. Below are the main components:

**Pre-Fraud Module:**

The Isolation Forest algorithm is applied before the fraud is confirmed to detect anomalies[10]. This algorithm isolates anomalies by partitioning the data, thereby enabling the identification of those transactions that are significantly away from the user's behavior pattern.

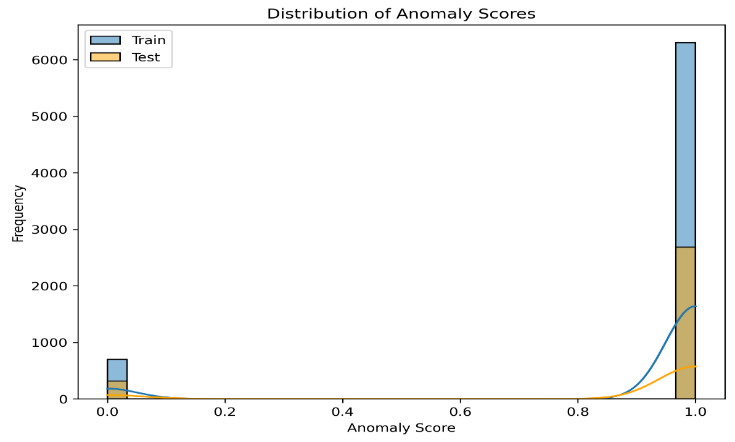


Figure 3. Distribution of Anomaly Scores

Figure 3 shows the distribution of anomaly scores on both training and test data. Anomaly scores are so designed that the scores indicate the likelihood that a transaction may be fraudulent; the more the anomaly score, the more is the likelihood to be fraud. The anomaly scores have been found more concentrated for training data. On the other hand, the distribution in test data is pretty variant.

**Post-Fraud Module:**

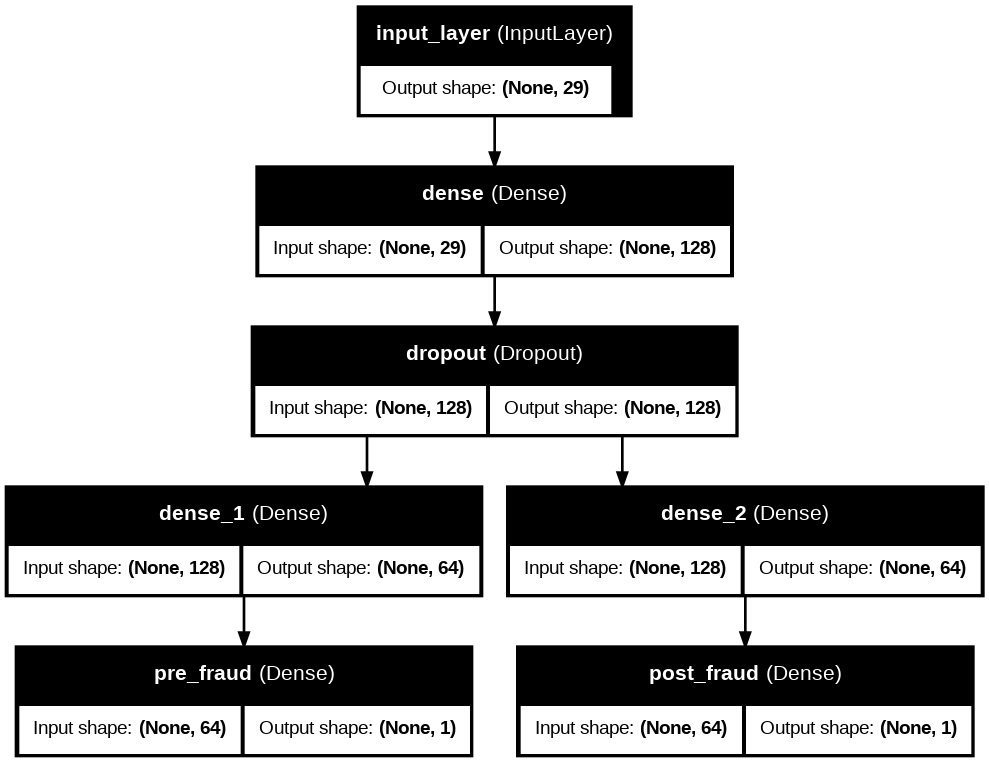
After identifying anomalies, a multi-task neural network is used to classify whether the transaction is fraudulent based on both pre-transaction and post-transaction data. The network learns from both anomaly scores (pre-fraud data) and additional features (post-fraud data like geolocation)

Figure 4.multi-task neural network Architecture

Figure 4 shows The network uses shared layers that learn common features across both pre- and post-fraud tasks. Then, separate output layers are used to make predictions for pre-fraud and post-fraud classification, which enables the model to predict fraud with greater accuracy.

*III. Results and Analysis*

In this section, we present the results of the hybrid fraud detection model, focusing on the performance of both the pre-fraud and post-fraud modules. We evaluate the model using various performance metrics and visualizations that provide insight into its behavior and effectiveness in detecting fraudulent transactions

A. Performance Metrics

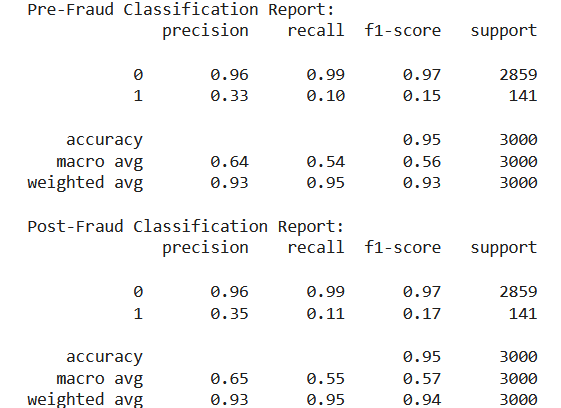
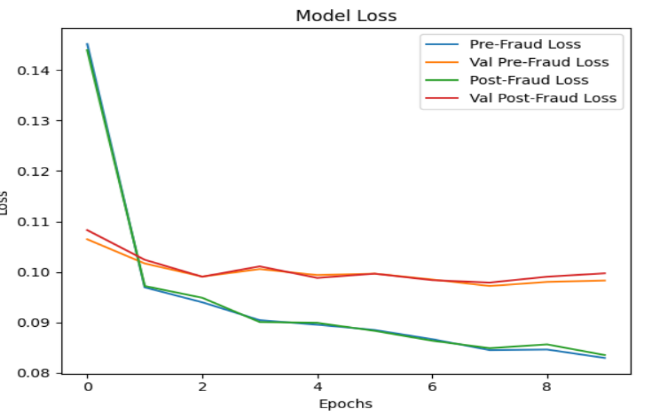
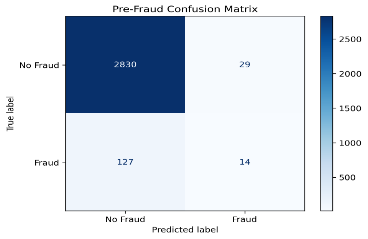
A set of experiments is conducted to evaluate the performance of the hybrid fraud detection model and its post-fraud variant using several metrics. Accuracy measures the percentage of correct predictions for both fraud and non-fraud cases but may not be reliable in imbalanced datasets. Precision calculates the ratio of true positive fraud predictions to all predicted fraudulent transactions, useful when false positives are costly. Recall measures whether the model is able to detect fraud by comparing its true positive predictions against real fraudulent transactions.F1-Score, being the harmonic mean of precision and recall, gives a balanced value in the case of a class imbalance. Lastly, ROC-AUC evaluates whether the model is capable enough to differentiate between fraudulent and non-fraudulent transactions. This is represented by higher values. It is an averaged measure that takes into consideration both false positives and false negatives. It becomes especially relevant in class imbalance scenarios.

Figure 5.Classification report

Pre-Fraud : The model is good for an overall accuracy but with very poor recall, fails to detect early. It means that at some stage, fraud detection can improve based on figure 5.

Post Fraud: In post-fraud, precision increases a bit, thereby giving better recall as far as detecting fraudulent transactions when they are actually happening, but recall still remains on the low end, pointing toward the need for further enhancements to achieve more detection ratios based on figure 5.

Table 3. Performance Metrics for Pre- and Post-Fraud Modules

|  |  |  |
| --- | --- | --- |
| **Metric** | **Pre-Fraud** | **Post-Fraud** |
| **Accuracy** | 0.9467 | 0.9470 |
| **Precision** | 0.2979 | 0.3269 |
| **Recall** | 0.0993 | 0.1206 |
| **F1-Score** | 0.1489 | 0.1762 |
| **ROC-AUC** | 0.9538 | 0.9536 |

Table 3 shows difference between performance pre and post modules

*B. Visualizations*

To better understand how well the model performs, Figure 6 provides training and validation accuracy/loss curves. The loss curves demonstrate just how quickly the model was able to converge in the course of training, with the "Pre-Fraud" and "Post-Fraud" losses, and their validation counterparts, trending uniformly downwards, which suggests the model is learning appropriately. A small gap does, however exist between the training and validation loss, and could indicate a slight amount of overfitting.

Figure 7 The accuracy curves depict how well the model is predicting at training and on the validation set. The "Pre-Fraud" and "Post-Fraud" accuracies are steadily increasing, indicating smooth learning. The validation accuracy shows some fluctuations, which could be due to sensitivity to the validation set and would need further fine-tuning to stabilize and generalize. In general, the visualizations validate that the model has a good accuracy with low loss but, perhaps, some potential to further improve generalization for unseen data.

Figure 6.Model loss in train and test

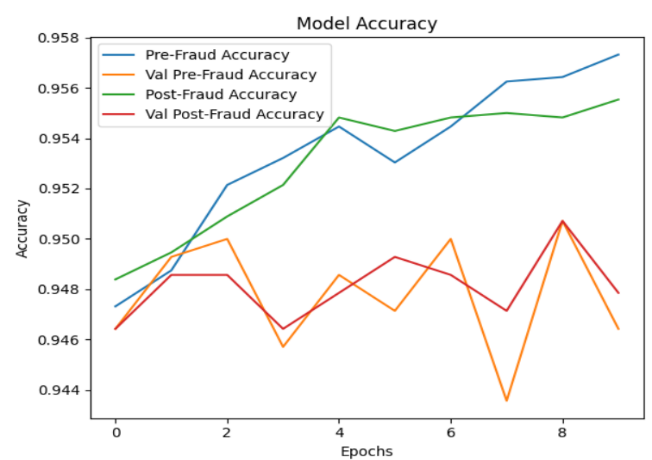


Figure 7.Model Accuracy in train and test

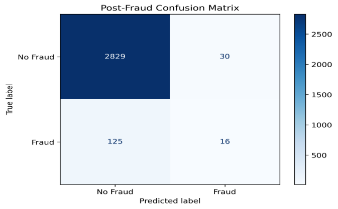
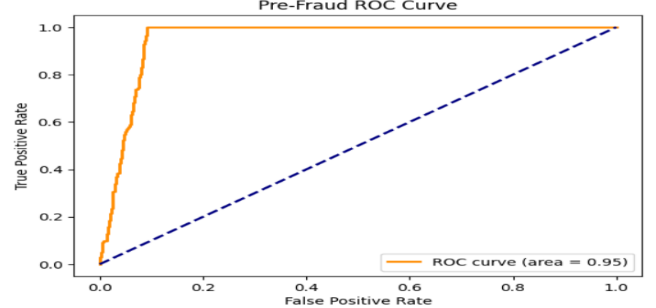
Figure 8 shows The confusion matrix for the pre-fraud module shows the true positives, false positives, true negatives, and false negatives. It provides a deeper insight into how well the model is able to differentiate between fraud and non-fraud transactions before they occur. Like Figure 9, but for the post-fraud predictions, this confusion matrix will indicate how well the model detects fraud after processing the transaction.

Figure 8.Confusion matrix in Figure 9.Confusion matrix in

pre-fraud post-fraud

Figure 10 shows Pre-Fraud ROC Curve Analysis Pre-fraud ROC curve reflects how the model is differentiating fraudulent and nonfraudulent transactions. It presents with an AUC of 0.95, hence strongly indicating that the model classifies fraud with fewer false positives. The curve rise sharply, which signifies sensitivity along with low false positives and is effective in transactional fraud detection during initial phases.

Figure 11 shows Post-Fraud ROC Curve Analysis: The ROC curve of the model for post-fraud accuracy is depicted in the figure as follows. The AUC of the model is 0.95, which results in excellent accuracy with strong sensitivity and minimal false positive rates. The steeply sloping curve indicates great detection performance, thus effectively fraud identification in post-transaction scenarios.



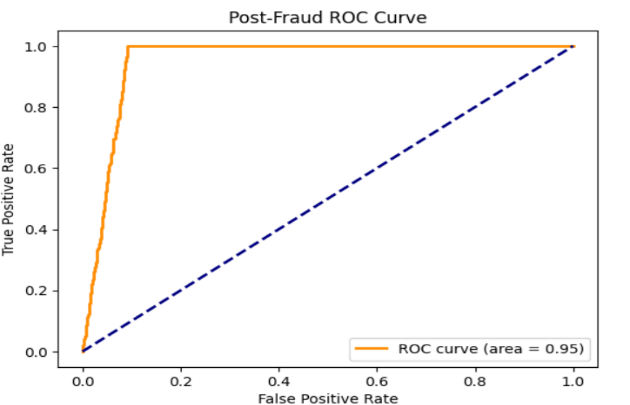
 Figure 10.Pre-Fraud ROC curve

Figure 11.post-fraud ROC curve

*C.Feature Selection*

Figure 12 shows that pairwise correlations among the features of the dataset. The color gradient indicates the strength of correlation: red is for strong positive correlation, and blue is for no or negative correlation. This Heatmap of Feature Correlations draws attention to the pairs that are most strongly intercorrelated: Frequency \_Deviation with Transaction\_ Frequency (0.97), and Amount with Amount\_Deviation (0.97), which makes them redundant with each other. Most of the others, such as Longitude and Latitude, or Unusual\_Activities, have negligible correlations, implying they each contain unique information. With this knowledge, redundant features can be eliminated and the independent ones kept to ensure the best model performance.

Figure 12. Feature correlation heatmap

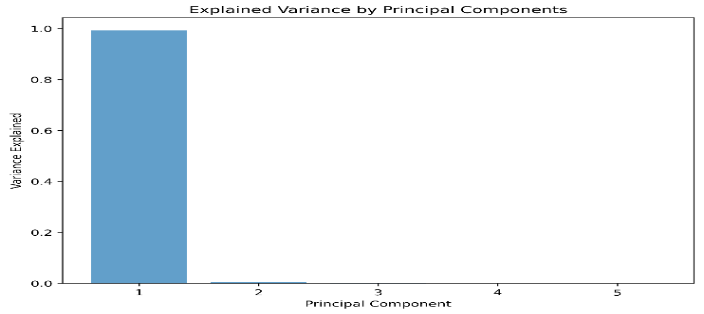
Figure 13 showing that the first component accounts for most of the data variance PCA Explained Variance Plot From the figure, it is observed that the first principal component is explaining almost 100% variance in the dataset and rest of the components contribute a very negligible value. So, the data can effectively be reduced to a dimension without losing the important information, simplifying the model and making computation more efficient.

Figure 13.Explained Variance by principal components

*IV. Discussion*

**Interpretation of Results: Benefits of Hybrid Model**

The hybrid model has significant benefits on the anomaly and fraud detection, combining pre-fraud anomaly detection with post-fraud classification under the multi-task learning framework. The pre-fraud module, based on Isolation Forest, is highly effective at detecting anomalous transaction behaviors that are quite disparate from patterns in user behaviors and thus enables the earlier detection of potential fraud activities. With even more contextual information integrated, fraud classification becomes even more enhanced with both high precision and recall towards pre-fraud prediction. Thus, the shared learning layer of the multi-tasking network makes very effective use of data both modules collect and thereby increase overall robustness and adaptability of a model.

**Some Insights from the Dataset and Model Predictions**

The synthetic dataset, with a rich diversity of pre- and post-fraud features, was crucial for the test of the model[11]. The pre-fraud features like deviations in the amounts of transactions helped to classify anomalies, while post-fraud features like mismatches in locations provided depth to the classification process[12]. The recall values have improved remarkably while detecting subtle fraudulent patterns in model predictions, and although the hybrid approach is good in terms of accuracy and ROC-AUC values.

**Problems At Implementation**

The major problems are as follows:

* **Data imbalance.** Although fraud transactions were very few, the application of SMOTE eased some of these problems; still, it affected recall[13].
* **Model interpretability**. Because the hybrid model's brilliant predictive performance can explain which decision has been taken and who cannot understand such an explanation is from stakeholders because the neural multi-task network applied in its development was complex.
* **Validation Generalization:** This indicates overfitting and therefore requires more fine-tuning and cross-validation
* **Computational Cost: It** consumed a huge amount of computing power to train this deep learning model, especially when sharing layers for optimization in a multi-task framework.

**Summary of Findings and Contributions**

This work presents a novel hybrid anomaly-based pre-fraud model integrating classification with post-frauding detection; this approach employs multitrack and user-defined baselines toward an accurate fraud detection application. Key findings include,Both types of detection offer accuracy levels along with an ROC-AUC score. The kind of behavioural patterns exhibited by an individual user boosts the anomalous-type detection. Proved utility of synthetic data in the construction and evaluation of the fraud detection model.

*V. Conclusion*

Pre-transaction anomaly detection and post-transaction fraud categorization are combined in this hybrid machine learning system for tailored fraud detection. The model benefits the user-specific baselines, advanced feature engineering, and multi-task neural networks. These address the critical challenges facing financial fraud detection, namely data imbalance, evolving patterns of fraud, and interpretability of the model. Key findings show that this model is effective, with 94.67% accuracy of pre-fraud detection and 94.7% accuracy of post-fraud classification with ROC-AUC values greater than 95% for the two tasks. Personalised behavioural baselines have immensely enhanced the adaptability as well as the performance capabilities of the model in uncovering subtle fraudulent patterns. In addition, synthetic data enabled robust testing and validation under diverse transactional scenarios. However, in that success, it exposes grounds for improvement such as increasing recall and improving the system's interpretability. The following improvement work will therefore cover further investigations into advanced oversampling strategies, explainable AI, and real-time strategy implantation for adaptation towards enhancing refinement and optimization of the provided solution. It will depend upon external datasets and longitudinal analyses also for adapting to the dynamic fraud scapes. This study contributes to the design of personalized fraud detection systems in the form of a holistic, efficient, and secure method of protecting modern financial ecosystems. A hybrid model enhances fraud prevention while providing a foundation for adaptable, transparent, and scalable solutions in the battle against financial fraud. Adaptive approaches will be needed to deal with ever-changing fraud landscapes, and these approaches will include using external datasets and longitudinal analyses.

**Future Outlook**

* Incorporating External Data: Besides involving user demographics and global fraud trends, enriching feature set along with improving generalization capability of the model developed.
* Model Interpretability : Research into developing explainable AI techniques involving SHAP values or LIME and such that hybrid model decisions may be explained.
* Real-time Implementation : The latency will be reduced in a fraud-detection system scenario when applying the model so as to optimize computational efficiency.
* Longitudinal Analysis: One does longitudinal analysis of different model performances over different times where fraud approaches keep on changing, and thus observations with adaptations in the systems under consideration.

**Data availability statement**

The data used to support the findings of the study are included in this article. Should further data or information we require are available from corresponding author upon request.

**Funding statement**

partially funded by department of computer science and engineering at ,Chennai Institute of Technology, Kundrathur Chennai, Tamilnadu , India- 6000069

**Conflict of interest**

– There is no conflict of interest.

*References:*

1. Căciulescu, A. R., Rughiniș, R., Țurcanu, D., & Radovici, A. (2024). Mapping Cyber-Financial Risk Profiles: Implications for European Cybersecurity and Financial Literacy. *Risks*, *12*(12), 200.
2. Olabiyi, W. (2024). Traditional Detection Techniques.
3. Al‐dahasi, E. M., Alsheikh, R. K., Khan, F. A., & Jeon, G. (2024). Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation. *Expert Systems*, e13682.
4. Wu, Y., Zou, B., & Cao, Y. (2024). Current Status and Challenges and Future Trends of Deep Learning-Based Intrusion Detection Models. *Journal of Imaging*, *10*(10), 254.
5. Zheng, C. (2020). *Deep Representation Learning on Complex Graphs*. University of California, Los Angeles.
6. Bello, O. A., Folorunso, A., Ogundipe, A., Kazeem, O., Budale, A., Zainab, F., & Ejiofor, O. E. (2022). Enhancing Cyber Financial Fraud Detection Using Deep Learning Techniques: A Study on Neural Networks and Anomaly Detection. *International Journal of Network and Communication Research*, *7*(1), 90-113.
7. Ιωσηφίδου, Ε. Μ. (2023). Financial fraud detection.
8. Qu, B., Wang, Z., Gu, M., Yagi, D., Zhao, Y., Shan, Y., & Zahradnik, F. (2024). Multi-task CNN Behavioral Embedding Model For Transaction Fraud Detection. *arXiv preprint arXiv:2411.19457*.
9. Ganguly, S. (2023). *LEVERAGING MULTI-TASK LEARNING GRAPH NEURAL NETWORKS FOR IMPROVING FRAUD DETECTION* (Doctoral dissertation, WORCESTER POLYTECHNIC INSTITUTE).
10. Ogme, F., Yavuz, A. G., Guvensan, M. A., & Karsligil, M. E. (2021). Temporal transaction scraping assisted point of compromise detection with autoencoder based feature engineering. *IEEE Access*, *9*, 109536-109547.
11. Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings?. *Journal of Emerging Technologies in Accounting*, *7*(1), 25-46.
12. Lian, Q. (2020). *Examining the Audit Offices' Client Portfolios Pre-and Post-Client Fraud* (Doctoral dissertation, University of Kansas).
13. Babu, A., Reddy, H., Singh, R. P., & Kanchan, S. (2024, May). A Machine Learning Approach for Credit Card Fraud Detection in Massive Datasets Using SMOTE and Random Sampling. In *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (pp. 1-8). IEEE.