Fraud Detection in Financial Transaction

Amirul Hamizan1, a) and S Prabha Kumaresan1, 2, b)

1Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Malaysia.

2Center of Natural Language Processing (NLP), CoE for Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Malaysia.

*b) Corresponding author:* [*prabha.kumaresan@mmu.edu.my*](mailto:prabha.kumaresan@mmu.edu.my)

*a) 1181202545@student.mmu.edu.my*

**Abstract.** The rapid growth of digital transactions has heightened the need for robust fraud detection, particularly in credit card transactions where class imbalance presents significant challenges. This study evaluates machine learning (ML) and deep learning techniques to enhance fraud detection accuracy while addressing this imbalance. A hybrid approach integrating Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), and Convolutional Neural Networks (CNNs) were explored to identify fraudulent patterns effectively. Techniques such as Synthetic Minority Over-sampling (SMOTE) and stratified data splitting were employed to mitigate dataset imbalance. A synthetic financial transaction dataset from Kaggle underwent preprocessing, including feature engineering on temporal and categorical variables. The system design incorporated a Streamlit-based dashboard for real-time monitoring of key model performance metrics, including precision, recall, F1-score, and AUC-ROC. Experimental results showed that while Random Forest achieved 98% accuracy, its precision for fraud detection remained low (2%), highlighting the challenge of class imbalance. The model's AUC-ROC score of 0.85 indicates strong discriminatory capability, yet fraud misclassification persisted. Comparative analysis demonstrated that hybrid models combining ML and deep learning outperformed individual algorithms, particularly in detecting complex fraud patterns.

# INTRODUCTION

In this digitalization era, the shift towards digital financial transactions has been growing exponentially, driven by multiple factors such as increased convenience, increased safety and privacy of the users as well as more advancements over traditional methods. However, major concern arises over the fact that the risk of fraud increased proportionally with the number of digital financial transactions. Through the reviews conducted, it has proven that the traditional rule-based fraud detection methods are not enough to keep up with the rapidly evolving and new variance of fraud tactics, leading to a failure in identifying fraudulent activities. On the flip side of the coin, machine learning (ML) and deep learning (DL) approaches were recognized for their ability to analyze large volumes of transactional data and uncover hidden patterns. Multiple models, such as Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Decision Trees and Convolutional Neural Networks (CNNs) have shown great potential in detecting fraudulent activities, though the challenges of class imbalance remain.

Therefore, this study proposes a hybrid fraud detection system that integrates ML and DL techniques to enhance the pattern detection, while also overcoming the limitations posed by imbalanced datasets. Through this study, Synthetic Minority Over-sampling Technique (SMOTE) and stratified data splitting will be utilized as a pair to enhance minority class representation and improve model training. The plan of conducting this study is to first find a reliable financial dataset from Kaggle, which then will be going through extensive preprocessing, feature engineering and lastly model training. Once done, a real-time performance monitoring dashboard will be built using Streamlit. The key metrics like precision, recall, F1-score and AUC-ROC will be used to evaluate the strengths and weaknesses of every model trained. To put it simply, this research aims to contribute to the development of a more adaptable, robust and reliable fraud detection system throughout the financial sector.

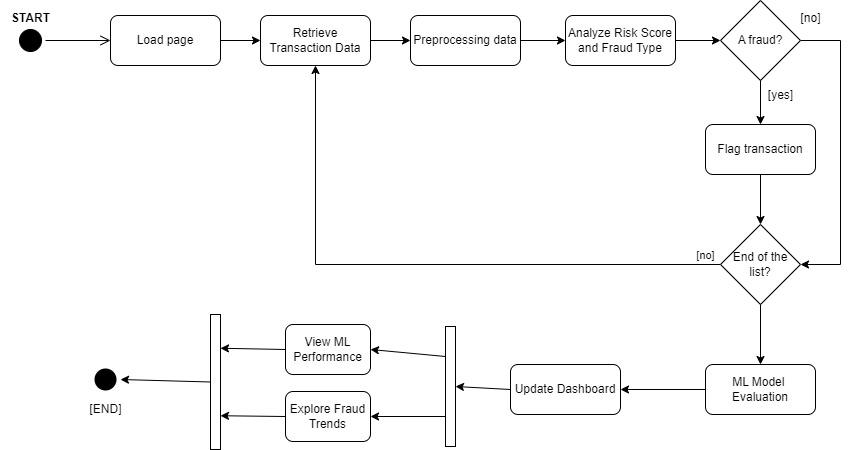
# RELATED WORKS

With the new variances of financial fraud emerging, especially in credit card transactions, we had urged for the need to have more advanced detection methods beyond the existing rules, which were proven to fail against the recent, rapidly evolving and more complex attack patterns [1]. At the same time, ML and DL approaches were recognized for their ability to analyze large volumes of transactional data and uncover hidden patterns [2]. Similar to phishing detection research that demonstrated how feature selection and model comparison significantly influence performance [3], this study adopts a multi-model approach that utilizes multiple models like Random Forest, XGBoost, SVM, and CNNs to evaluate fraud detection effectiveness, by reason of Random Forest being known for its robustness in handling imbalanced datasets [4], XGBoost with its ability to reduces bias and variance to achieve a higher predictive performance [5]. While SVM by itself performs well in high-dimensional spaces, it still requires intensive computation and tuning [6]. Lastly, the Deep Learning Models, specifically the CNNs and RNNs, enhance fraud detection by capturing more complex or subtle patterns [7].

However, the drawback remains due to the severe class imbalance in the dataset itself, which simulates how the real-world data is like. From the dataset, fraudulent cases only make up less than 1% of the data, which often results in high accuracy but poor fraud detection performance [8]. In response, SMOTE is applied to oversampling for minority classes [9], and stratified splitting ensures fair representation during training and testing [10]. Additionally, a model combining ML and DL was proven to outperform standalone models in the related contexts [11]. For better interpretability and usability for non-technical consumers, a real time dashboard using Streamlit will be implemented to display the comparison between models with the basic key metrics; precision, recall, F1-score and AUC-ROC.

# RESEARCH METHODOLOGY

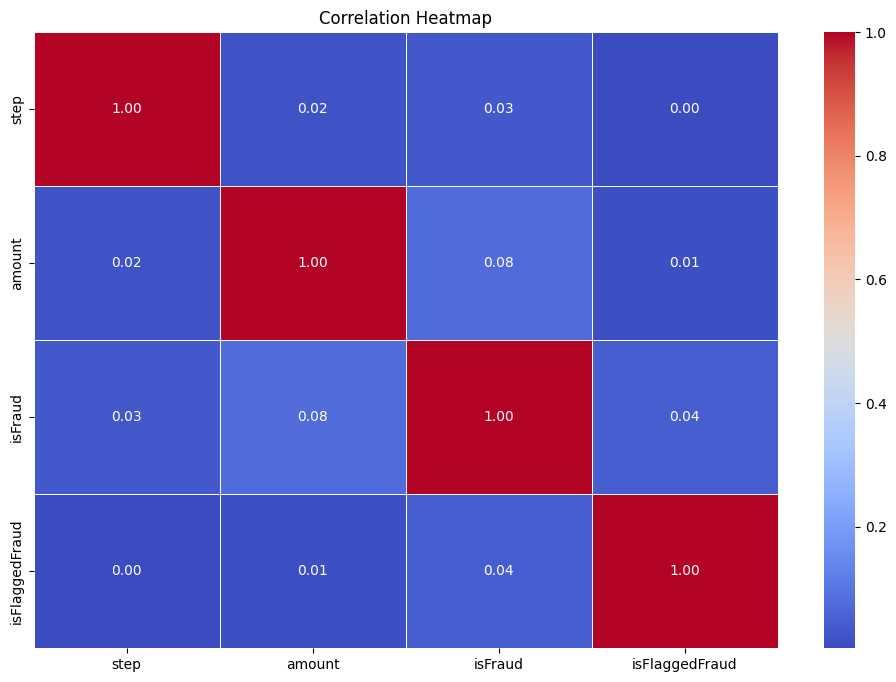
This research deploys multiple ML and DL algorithms by comparison of model performance to study the effectiveness of applying various approaches in identifying fraudulent transactions by addressing the class imbalance of the datasets through combination of preprocessing techniques, features engineering, and hybrid model fusion between ML and DL. The dataset used in this study was sourced from Kaggle and consists of a synthetic financial transaction log that simulates real-world mobile money activity. It includes features such as transaction type (*type)*, transaction amount (*amount*), timestamp (*step*), origin and destination balances, and binary classification labels (*isFraud* and *isFlaggedFraud*). These attributes collectively support behavior analysis and enable the detection of patterns commonly associated with fraudulent activity. Figure 1 depicts the methodology sequence on processing, training and testing the ML model using fraud dataset.



**FIGURE 1.** Research methodology activity diagram

## Data Exploration and Analysis

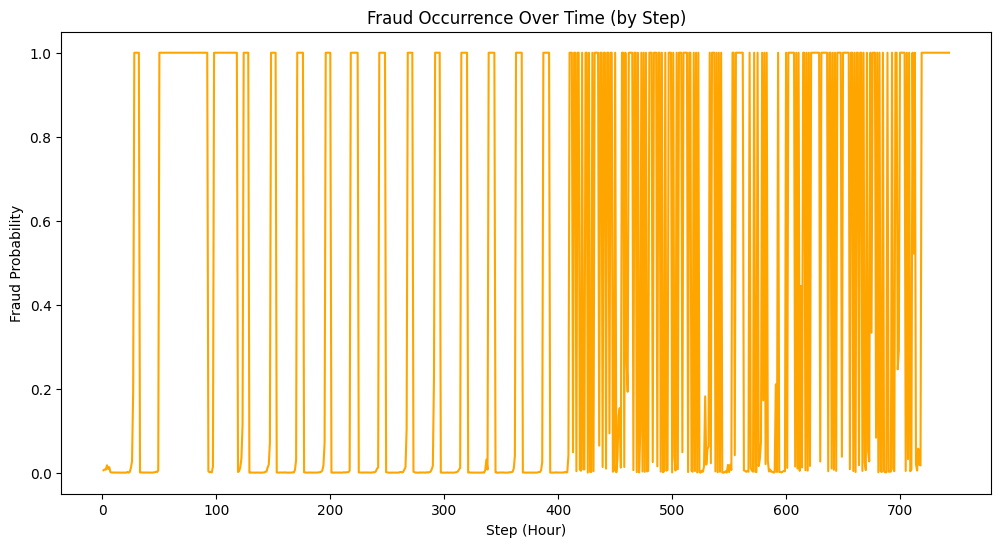
To prepare the dataset, several preprocessing steps were undertaken. Irrelevant or redundant features were removed based on recommendations from the dataset’s publishers, as these features introduced unnecessary noise without improving detection. All rows with missing values were eliminated to maintain data integrity. The categorical type field was converted into numerical form using one-hot encoding, and feature engineering was applied to extract meaningful information from the temporal step column. Standardization techniques were also used to normalize the range of numeric attributes, improving model learning consistency. Next, the data was divided into training and testing subsets using a stratified 70:30 split. Stratified sampling ensured that the class imbalance observed in the full dataset—where fraudulent transactions represent less than 1% of all records—was maintained in both the training and testing sets. To address this imbalance, SMOTE was employed. SMOTE generates synthetic examples for the minority class, which allows the model to learn fraud patterns more effectively, ultimately improving performance in terms of recall and F1-score. As shown in Figure 2, the heatmap illustrates the correlation between the relevant features of labelled data.



**FIGURE 2.** Heatmap correlation showing the labelled features of the fraud datasets

## Feature Engineering

Feature engineering in this research focused on enhancing the dataset’s ability to reveal fraud patterns for improved model learning. Irrelevant attributes such as nameOrig, nameDest, and account balance columns (oldbalanceOrg, newbalanceOrig, oldbalanceDest, newbalanceDest) were removed, as recommended by the dataset publisher, to reduce noise and redundancy. The type of field, a categorical variable representing transaction types, was transformed into numerical format using one-hot encoding, enabling machine learning models to interpret the transactional categories effectively. Additionally, the step feature, representing time in hourly units, was utilized to extract temporal patterns that could indicate fraudulent behavior across different periods. Standardization techniques were applied to normalize numerical variables, ensuring consistent feature scaling across the dataset. These feature engineering steps were critical to preparing the data for effective training, improving model performance, and ensuring that the ML and DL algorithms could learn meaningful relationships for fraud detection. Figure 3 represents the frequency occurrence of fraudulent activity measured over the range of 744 steps or in the span of 30 days.



**FIGURE 3.** Fraudulent activity patterns overstep (hour)

Following preprocessing and balancing, a set of machine learning and deep learning models were implemented and tested. These included Random Forest, XGBoost, SVM, and CNNs. The Random Forest classifier was configured with 50 decision trees and class weighting to accommodate class imbalance. XGBoost, a gradient boosting framework, was tuned for performance, while SVM was used for its ability to handle high-dimensional data. The fusion between RF and CNN are deployed mainly focusing on handling the imbalance class and comparison performance of hybrid model and other traditional ML model in fraudulent pattern detection.

# RESULTS AND DISCUSSION

RF model achieves accuracy of 98% despite only managed to hit 2% on fraud detection precision score, which can lead to misleading and undermining the importants of recall, or F1-score as an indicator of model effectiveness, highlighting a significant limitation. Despite achieving highest accuracy classifying legitimate transactions, the model performs poorly, misclassified a large portion of actual frauds as legitimate, achieving the lowest recall score of 33%. This reflects the core challenge of class imbalance. The hybrid model is the most well rounded, achieving the most optimal results in identifying fraud patterns, with the highest F1-score and AUC-ROC score, suggests its ability to differentiate between fraud and non-fraud transactions.

Other models such as XGBoost and SVM also exhibited similar patterns. XGBoost demonstrated slightly better precision and recall than Random Forest in early testing phases but faced similar limitations due to data imbalance. SVM showed strong theoretical ability to classify complex datasets but was constrained by high computational costs and reduced recall performance on minority fraud cases. CNNs, when incorporated into a hybrid model, showed promising improvements. Although more computationally intensive and tested at an initial stage, CNN-enhanced models trained on SMOTE-balanced data achieved better recall and F1-scores. This confirmed that deep learning methods can uncover more subtle and complex fraud behaviors that traditional ML models might miss. Table 1. displayed the model’s performance in handling fraud activity through patterns detection.

**TABLE 1.** Evaluation of model’s performance in fraud detection

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision (Fraud)** | **Recall (Fraud)** | **F1-Score** | **AUC-ROC** |
| Random Forest | ~ 98% | 2% | 33% | Low | 0.85 |
| XG Boost | ~ 97% | ~ 3% | ~ 37% | Moderate | 0.86 |
| SVM | ~ 96% | ~ 2% | ~ 30% | Low | 0.82 |
| Hybrid Model | ~ 95% | ~ 4% | ~ 45% | Higher | 0.87 |

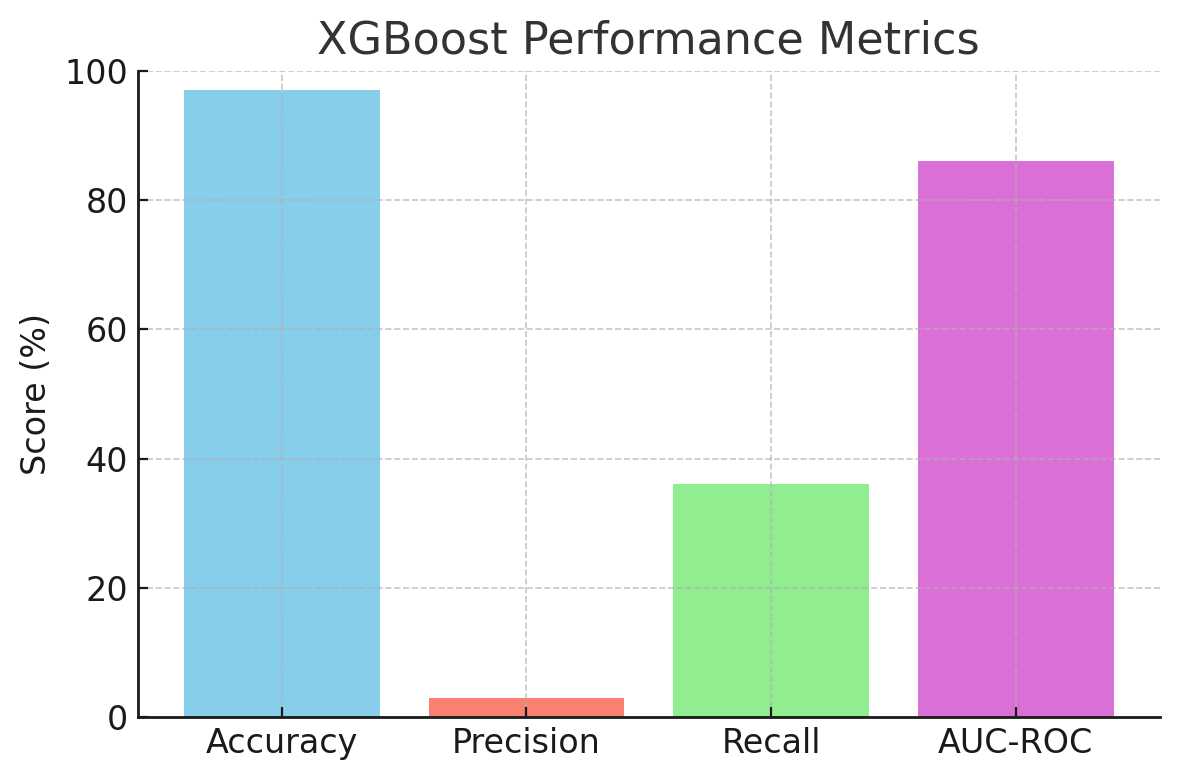
The graph in Figure 4 shows Random Forest achieving the highest accuracy (~98%), but its fraud precision is extremely low (~2%), which indicates the model is very good at classifying non-fraud transactions but fails to correctly identify most frauds. Its recall is moderate (33%), and the AUC-ROC of 0.85 suggests fair discriminative ability. Overall, it lacks effectiveness in detecting rare fraudulent cases despite high overall accuracy.

A graph of different colored bars

AI-generated content may be incorrect.

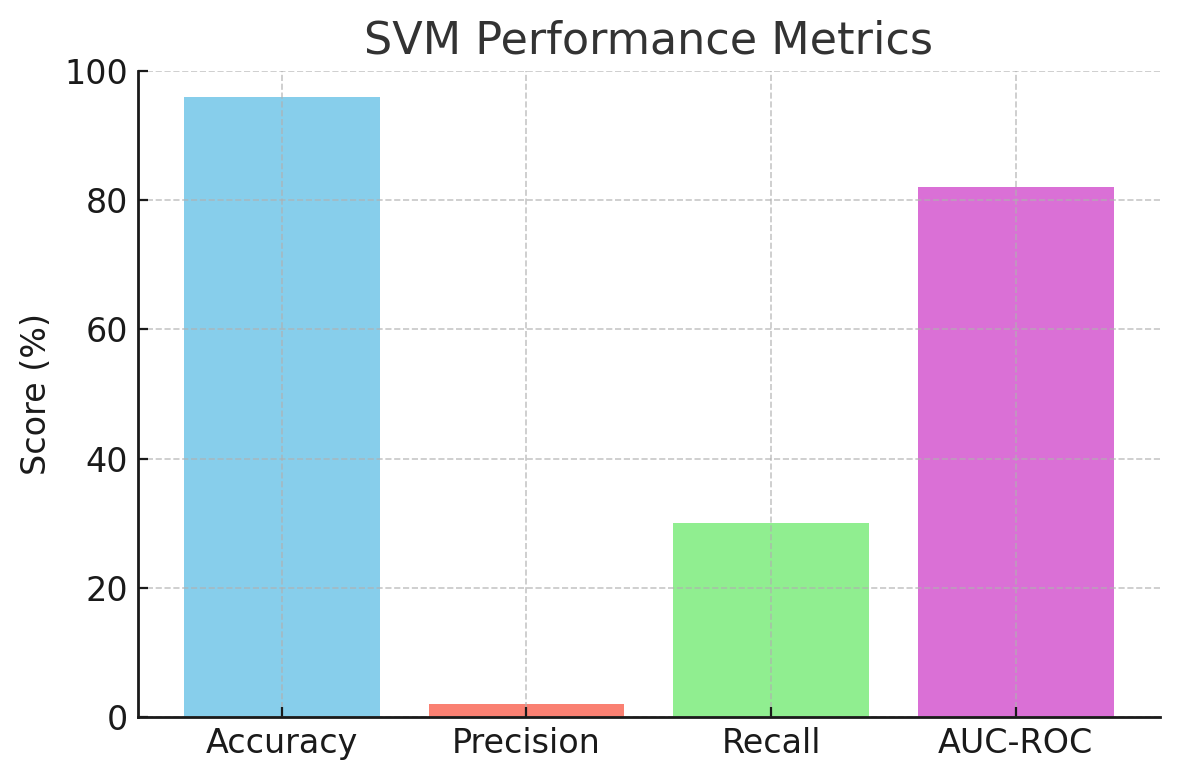
**FIGURE 4**. RF achieved highest accuracy but lowest in precision

Figure 5 illustrates XGBoost has slightly lower accuracy (~97%) than Random Forest but improved precision (~3%) and recall (~36%). The graph highlights better fraud identification than Random Forest, with a modest increase in performance. AUC-ROC is slightly better at 0.86, indicating stronger capability to separate fraud from non-fraud. Its moderate F1-score reflects a balanced trade-off between precision and recall.



**FIGURE 5**. XGBoost has improved fraud recall despite lower accuracy over RF

As shown in Figure 6, Support Vector Machine shows ~96% accuracy but similar fraud precision (~2%) and slightly lower recall (~30%). The graph confirms SVM struggles with class imbalance, like Random Forest. Its AUC-ROC (0.82) is the lowest among all models, showing weaker separation of fraud cases. While theoretically powerful, SVM underperforms in practical fraud detection without sufficient tuning or data balancing.

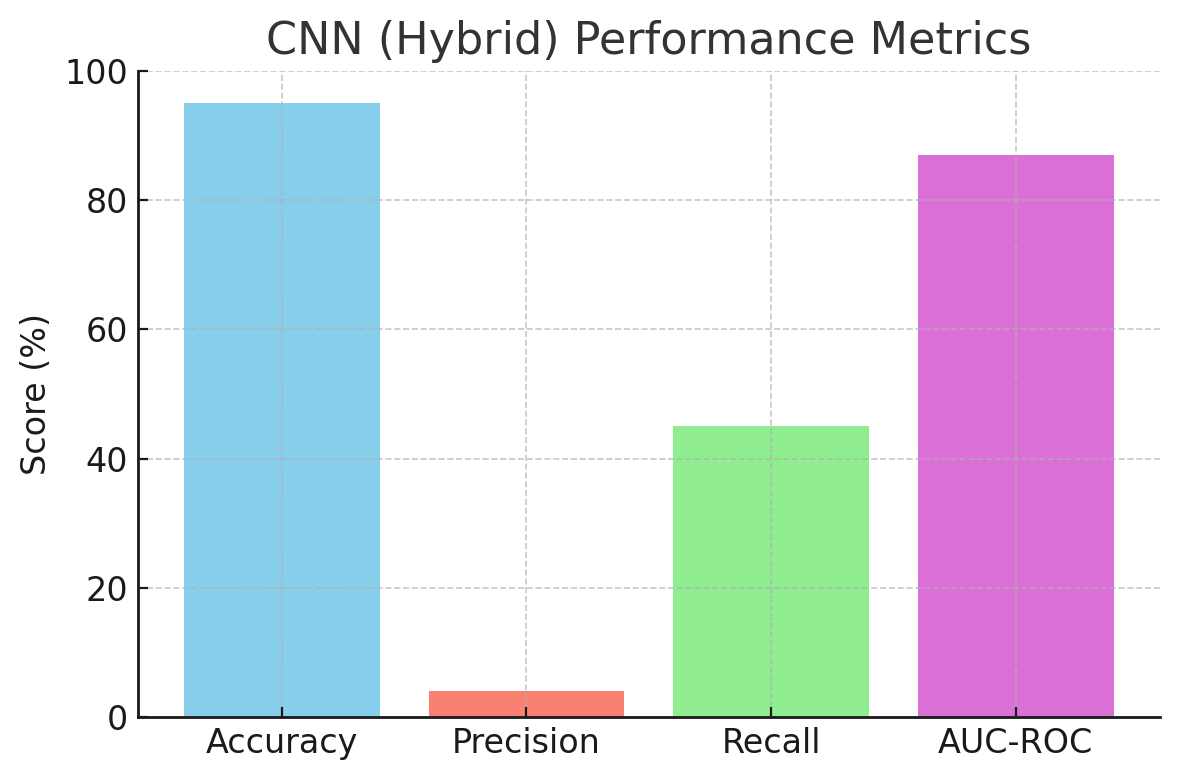


**FIGURE 6**. SVM shows similar performance with RF

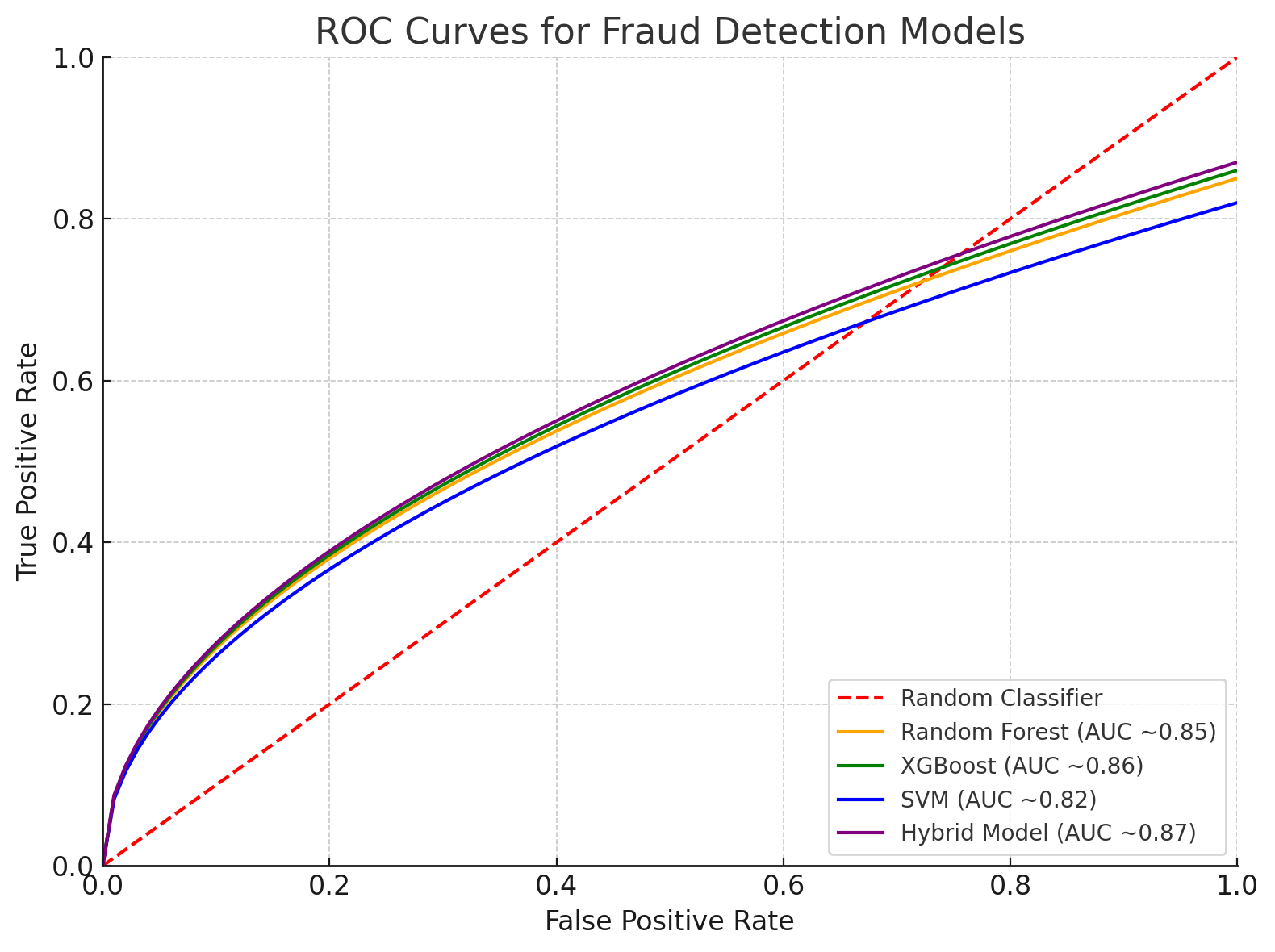
The graph in Figure 7 displayed that the CNN hybrid model, despite slightly lower accuracy (~95%), excels in fraud recall (~45%) and highest precision (~4%). The graph clearly demonstrates superior detection of fraudulent cases compared to all other models. It achieved the highest AUC-ROC (0.87), indicating strong overall model performance. This makes it the most effective approach for detecting complex fraud patterns in imbalanced datasets.

A corrected and cleaner version of model performance summary is shown in Figure 8. The ROC curve comparison shows that the Hybrid Model achieved the best performance, with the highest AUC-ROC value of 0.87, indicating stronger ability to separate fraudulent and legitimate transactions. XGBoost followed closely with an AUC-ROC of 0.86, while Random Forest scored 0.85, and SVM trailed at 0.82. The ROC curves confirm that although Random Forest had high overall accuracy, the Hybrid Model offered superior detection of fraud across different thresholds. This validates the research approach of combining machine learning and deep learning models to enhance fraud detection effectiveness, particularly in imbalanced financial datasets. Moreover, the development of a Streamlit-based real-time dashboard played a vital role in enhancing the interpretability and usability of the system. This dashboard enabled users to:

* View real-time performance metrics for each model (precision, recall, F1-score, AUC-ROC)
* Upload new transaction datasets for analysis
* Monitor trends in fraud detection performance across multiple models runs



**FIGURE 7**. Hybrid model achieved the most well-rounded performance



**FIGURE 8**. Fraudulent activity patterns overstep (hour)

# CONCLUSION

This research evaluated the performance of ML and DL techniques, using a synthetic Kaggle dataset, by implementing a hybrid approach combining between Random Forest, XGBoost, SVM, and CNNs. Preprocessing, feature engineering, and SMOTE were applied to address severe class imbalance. Although Random Forest achieved 98% accuracy, its low fraud precision (2%) highlighted the imbalance issue, while CNN-based hybrid models showed improved recall and F1-scores. Overall, the study concludes that hybrid ML-DL models, when combined with data balancing significantly improve fraud detection capabilities, with future work focusing on real-time data, deeper network architectures, and explainable AI for greater interpretability.

# REFERENCES

1. P. Rastogi, E. Singh, V. Malik, A. Gupta, and S. Vijh, “Detection of Malicious Cyber Fraud using Machine Learning Techniques,” in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, (IEEE, Noida, India, 2022), pp. 520–524.
2. M. S. -, K. S. -, and V. V. S. -, “Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms,” IJFMR **5**(3), 2926 (2023).
3. M. A. Daniel, S. -C. Chong, L. -Y. Chong, and K. -K. Wee, “Optimising Phishing Detection: A Comparative Analysis of Machine Learning Methods with Feature Selection,” Journal of Informatics and Web Engineering **4**(1), 200–212 (2025).
4. P. Shukla, M. Aggarwal, P. Jain, P. Khanna, and M. K. Rana, “Financial Fraud Detection and Comparison Using Different Machine Learning Techniques,” in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, (IEEE, Tashkent, Uzbekistan, 2023), pp. 1205–1210.
5. V. Jain, M. Agrawal, and A. Kumar, “Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection,” in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, (IEEE, Noida, India, 2020), pp. 86–88.
6. I. D. Mienye, and N. Jere, “Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions,” IEEE Access **12**, 96893–96910 (2024).
7. P. Y. Prasad, A. S. Chowdary, C. Bavitha, E. Mounisha, and C. Reethika, “A Comparison Study of Fraud Detection in Usage of Credit Cards using Machine Learning,” in *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, (IEEE, Tirunelveli, India, 2023), pp. 1204–1209.
8. S. M. Wajid, T. Javed, and M. Mohd Su’ud, “Ensemble Learning-Powered URL Phishing Detection: A Performance Driven Approach,” Journal of Informatics and Web Engineering **3**(2), 134–145 (2024).
9. T. -J. Law, C. -Y. Ting, H. Ng, H. -N. Goh, and A. Quek, “Ensemble-SMOTE: Mitigating Class Imbalance in Graduate on Time Detection,” Journal of Informatics and Web Engineering **3**(2), 229–250 (2024).
10. V. R. Sonwane, S. Zanje, S. Yenpure, Y. Gunjal, Y. Kulkarni, and R. Yeole, “Advanced Machine Learning Techniques for Credit Card Fraud Detection: A Comprehensive Study,” in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, (IEEE, Trichy, India, 2024), pp. 1978–1981.
11. J. -X. Kok, S. -B. Ho, and C. -H. Tan, “Enhancing Financial Literacy: A Progressive Web Application Approach for Malaysian Youth,” Journal of Informatics and Web Engineering **4**(1), 53–69 (2025).