International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

Int. J. Sci. Res. Arch.
IJSRA
International Journal of Science and Research Archive
Research Journal Archive, INDIA

(RESEARCH ARTICLE)

Intelligent Detection of E-Wallet Transaction Fraud Using Hybrid Deep Learning and Ensemble Machine Learning Models

Fhyruz Salsabil * and Abu Syed Md. Mostafizur Rahman

Department of Computer Science and Engineering, Bangladesh University of Professionals, Dhaka-1216, Bangladesh.

International Journal of Science and Research Archive, 2025, 17(01), 1156-1166

Publication history: Received on 19 September 2025; revised on 25 October 2025; accepted on 27 October 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.17.1.2923

Abstract

The fast development of the digital payment systems has created more danger of fraudulent transactions in e-wallet services. In this study, the proposed fraud detection system is an integrated system of machine learning and deep learning to improve the accuracy of fraud detection. The dataset used in developing models was a synthetic mobile money transaction dataset on Kaggle. Preprocessing of the data entailed the elimination of redundant attributes, label encoding and scaling the features with MinMaxScaler and selecting the features with Lasso regression. Several machine learning models were tried such as the Random Forest, K-Nearest Neighbors, Decision Tree, Logistic Regression, and XGBoost. An ensemble model was also implemented between RF and XGB. Moreover, the deep learning schemes such as Artificial Neural Network, Convolutional Neural Network, Recurrent Neural Network and a hybrid CNN+RNN model were created and optimized through hyperparameters. The results demonstrate superior performance from the CNN (92.24%) and hybrid CNN+RNN (92.22%) models, outperforming traditional machine learning approaches. This hybrid deep learning system offers a powerful and expandable solution to real-time fraud detection in mobile payment systems, which enhances safe and reliable digital financial systems.

Keywords: E-Wallet; Lasso; Machine Learning; Deep Learning; Ensemble Learning; Mobile Payment Security; Synthetic Dataset

1. Introduction

The digital revolution has had a significant change in the financial transactions of the global world since it allows payment services to be done easily and instantly via mobile wallet and other financial technologies that have been implemented online. Mobile wallet solutions have become central elements of the digital economy that provides users with a smooth payment experience, peer-to-peer transfers, and online buying processes. However, the growth of the digital payment systems also increases the susceptibility to the fraudulent activity. The anonymity, speed, and quantity of the electronic transactions provide favorable grounds upon which ill-intentioned actors exploit the vulnerability of systems leading to financial losses and lost consumer confidence. This has prompted the identification and mitigation of fraudulent transactions in e-wallet systems to emerge as one of the most critical issues in e-financial technology (FinTech) security.

The conventional rule-based fraud detection schemes are mainly relying on explicitly defined patterns and threshold parameters. Although these methods are simple and easy to interpret, they are often unable to identify complex and changing fraudulent behavior. Fraudsters frequently adapt their strategies to overcome these fixed limitations hence undermining the effectiveness of traditional systems. This has in turn forced researchers to explore data-driven, intelligent detection models that are able to adapt to the emergent threats and be able to learn through transactional behavior. In the context of this, machine learning and deep learning have received significant attention because of their ability to identify latent correlations in high-dimensional data volumes, as well as to identify valid and valid transactions.

* Corresponding author: Fhyruz Salsabil

Recent empirical studies have highlighted the effectiveness of machine-learning frameworks such as Random Forests, KNN, Decision Trees, Logistic Regression and XGBoost against detecting fraud transactions. However, all of the models have their own strengths and weaknesses that also depend on the inherent features of the dataset and the interactions among the features. Ensemble learning techniques have been used to overcome the constraints of individual algorithms by consolidating a number of classifiers and, therefore, improving predictive accuracy and performance. On the other hand, deep-learning models like the Artificial Neural Networks, the Convolutional Neural Networks and the Recurrent Neural Networks have been shown to be better at deco founding complex nonlinear relationships and sequential transactional dynamics. Deep-learning systems that combine CNN and RNNs, or any other hybrid approach, can take advantage of both spatial and time-related features, which can provide even greater fraud-detection capabilities.

This study aims to come up with a smart fraud detection system of e-wallet transactions by combining ensemble machine learning applications and blended deep learning architectures. Based on the synthetically obtained mobile money transaction dataset, which is available on Kaggle, the proposed study provides a detailed preprocessing procedure, such as data cleaning, feature scaling with the help of MinMaxScaler, label coding, feature selection with Lasso regression. Several machine learning and deep learning models are discovered and trained, optimized based on hyperparameter tuning, and tested based on their performance in classification. Comparison outcomes show that hybrid deep learning, which is CNN and CNN-RNN architecture, is better than traditional machine learning since the accuracy score goes beyond 92%.

The contribution of the current study is the fact that the hybrid deep learning and ensemble machine learning approaches can be applied in a synergistic way to ensure the creation of a powerful, scalable, and efficient system of detecting fraud in e-wallet systems. The fulfilled framework enhances the safety and dependability of online financial systems, which contributes to safe FinTech ecosystems and strengthens the trust of customers in mobile-based financial transactions.

2. Literature review

Jurgovsky et al. [1] presented the problem of fraud detection as sequence classification and indicated that recurrent models (LSTM) are effective in credit-card transaction sequences, asserting that the detection is enhanced by time-dependent information (prior transactions). They demonstrate that LSTM sequence models can detect repetitive trends of frauds in the short term that were previously missed by untrained models.

Carcillo et al. [2] explored streaming active-learning approaches and revealed that realistic fraud-detection systems should support labeling latency as well as investigator feedback. They suggested proactive criteria of selection on querying labels in streams hence demonstrating the importance of exploration exploitation trade-offs when investigations are constrained to a small number of transactions per day. This paper highlights the fact that when streaming and labeling constraints are not taken into account on offline evaluation protocols, the estimates of performance in a real-world context are likely to be overstated.

Dal et al. [3] provided a pioneer study of operational constraints realistic and relevant to large and long streams of transactions; this is in terms of class imbalance, concept drift, and verification latency. They proposed a learning approach and the related metrics of evaluation that are designed specifically to production fraud-detection systems and showed that traditional offline methods, like random train/test split and traditional performance measurements, are capable of making false inferences. Their approach has since been adopted as a standard on realistic experimental designs.

Whitrow et al. [4] studied the concept of transaction aggregation as a preprocessing strategy, in which recent activity of a cardholder is summarized into feature representations. The authors have shown that aggregation can improve the detection performance by reducing noise as well as providing more information about the behaviour. The article has been widely referred to as an example of how the choice of temporal aggregation windows can have a significant practical effect on the performance of classifiers.

Fiore et al. [5] performed an empirical study of the data augmentation method based on a generative adversarial network (GANs) to overcome the dramatic class imbalance in fraud detection. The GAN-based augmentation was able to make classifiers more sensitive to instances of fraud, and not cause any overfitting by synthesizing realistic minority-class samples even in the case where the labeled fraud instances are few. This direction of research has prompted many successive works that incorporate generative models together with discriminative detectors.

XGBoost developed by Chen and Guestrin [6] is not directly targeted at detecting fraud but in the industry, it is being used as a standard in detecting tabular fraud problems because it is robust, scalable and can support sparse features. Many comparative studies and practical fraud research use XGBoost as a comparison point or as part of an ensemble stack as it often achieves a great deal of accuracy at a reasonable level of interpretability using feature-importance scores in tabular financial data.

The scaling streaming architecture, built with Kafka, Spark, and Cassandra, that is made up of machine-learning strategies, by SCARFF (Carcillo et al. [7]), is capable of managing concept drift and class imbalance in near real-time applications. The SCARFF model is between the software engineering and algorithmic decision-making, which explains that scalable data pipelines and custom learning algorithms are required to successfully detect relevant fraud in a production-grade manner.

According to Nguyen et al. [8] and others, the importance of prospective information and time locality during detection of fraud has been highlighted and thus, most of the events of a fraud are represented in a short burst of transactions after a primary compromise thus hinting that a transaction that follows the indication of a suspicious activity is predictive. These findings suggest the implementing of sequence/historical/window schemes and designing of time characteristics.

There is a noticeable shift in recent academic literature dating between 2023 and 2025, with the focus on applying attention mechanisms, graph neural networks (GNNs), and hybrid models, as the means of capturing more complicated relational and time trends. Zhao et al. and various other works performed between 2024 and 2025 can assert substantial performance gains due to the use of self-attention mechanisms to model long-range interactions and GNNs to code transaction-entity graphs related to users, merchants, and devices; higher-order rings of fraudsters and orchestrated attacks that are inaccessible by single-transaction models can be identified. Additionally, these approaches often use the combination of attention or GNN-learned representations with convolutional or recurrent neural network features in stacked or fused model architectures. [9][10].

The methodological reviews and surveys indicate repeatedly that evaluations encounter the following issues: (a) pitfalls in the evaluation processes, such as inappropriate data split and inflated performance measures; (b) concept drift, which makes the use of cost-sensitive evaluation metrics essential, and online or adaptive learning methodologies may be required; and (c) the need of interpretability and explainability, which may be required to gain investigator acceptance. Also, various surveys support the combination of ensemble machine-learning algorithms, including the random forests and XGBoost, with deep-learning algorithms to learn sequences and representations, and then adding interpretability layers to enable operationally deployable results [11].

3. Methodology

The extensive discussion of the proposed methods will be introduced in this section. The general approach of the proposed E-Wallet Transaction Fraud Detection system is presented in Figure 1.

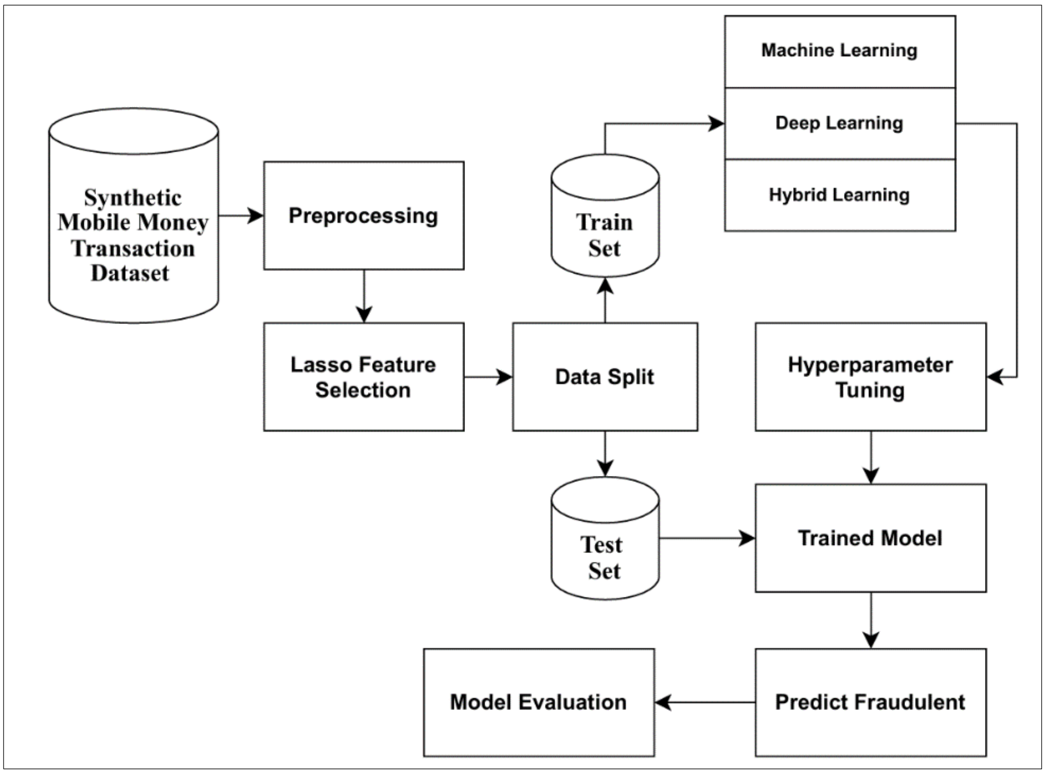


Figure 1 Proposed Methodology of This Study

3.1. Dataset Collection

The data used in the study was taken out of the publicly available Synthetic Mobile Money Transaction Dataset that is cited as [12] and is hosted on the Kaggle platform. This data has been designed to replicate actual-life mobile financial transactions and is based on the PaySim simulator which produces both valid and fraudulent transactions based on real financial transactions of a mobile money service provider. PaySim has been well known as having the potential to simulate dynamic behaviours of transactions without compromising privacy by anonymising sensitive data. Therefore, synthetic data would have statistical characteristics and behavioural patterns of real mobile transactions and thus make the suitable benchmark to be used in research on fraud detection. The data also includes a variety of transaction modalities, which are associated with a different operation in an e-wallet ecosystem. The transaction record attributes include but are not limited to: transaction amount, pre- and post-balance, source and destination account identifiers, transaction type and a binary label which shows whether it is a fraudulent transaction or not. All these attributes provide a complete description of user and system behaviour in mobile payment settings. In the current work, the data was obtained in the CSV format at the Kaggle and then loaded to a Python-based information processing system. The preliminary exploratory data analysis showed that there are a lot of transaction cases, and legitimate records significantly exceed fraudulent ones, which was the manifestation of the class imbalance characteristic of real-world financial datasets.

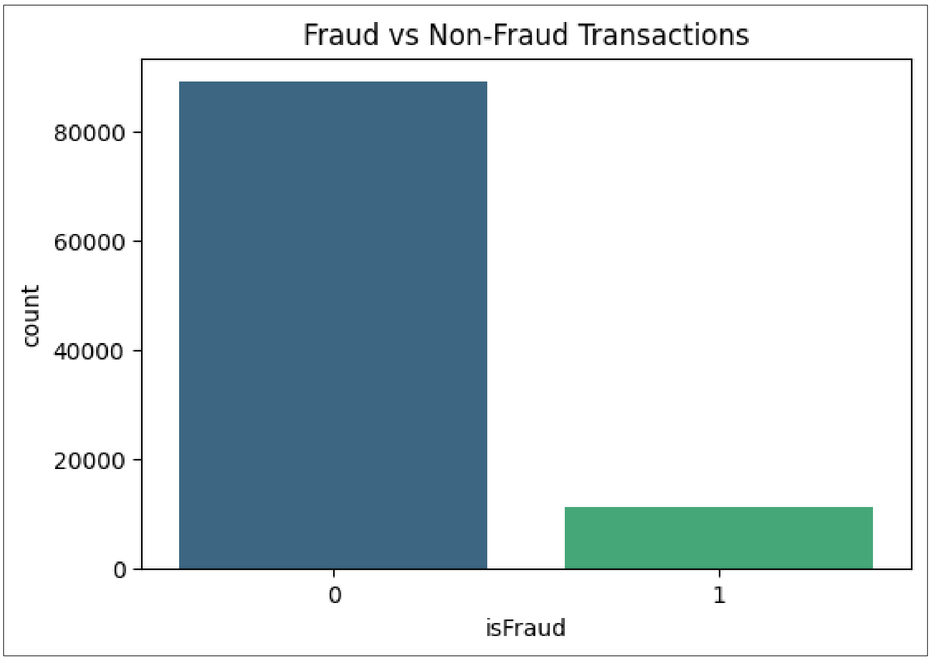


Figure 2 Data Distribution of SMMTD Dataset

3.2. Pre-processing

A set of preprocessing operations was done to assure both the integrity of the data and the preparedness of the model:

- *Cutting of Irrelevant Columns*: Irrelevant variables to the predictive modeling exercise like unique identifiers or redundant transaction flags were eliminated to decrease the extraneous noise and minimized the computational costs.
- *Missing values:* A thorough analysis of missing data was conducted, and data records which had missing or conflicting information were either imputed with the help of the relevant statistics techniques or excluded in the data.
- *Label encoding:* Categorical data were encoded into numbers using the LabelEncoder class, thus enabling them to work with subsequent machine-learning algorithms.
- *Feature scaling:* MinMax scaler was used to normalize continuous numerical variables so that all variables are within the same numerical range, which, in its turn, facilitates easier model convergence, as well as, the overall predictive performance.

3.3. Feature Extraction using LASSO

The extraction and choice of features is vital in the improvement of the model performance, interpretability, and computational efficiency, especially with a large dimensional transaction data. This paper has used Least Absolute Shrinkage and Selection Operator (LASSO) to determine the strongest predictors of e-wallet fraudulent transactions. LASSO is an effective regularization regression technique, which simultaneously carries out the task of feature selection and coefficient shrinkage. LASSO works by adding a L1 penalty term to the cost of a regression model making feature coefficients with less significant coefficients move towards zero. In the context of e-wallet fraud detection LASSO was used after the preprocessing occurred (MinMax scaling and label encoding). The aim was to reduce the number of most relevant transactional characteristics that have a strong impact on the likelihood of fraudulent activity. The model was also in a position to get rid of redundant or loosely related variables to focus on major predictors, such as transaction amount, balance differences, and indicators of transaction type.

3.4. Dataset Splitting

To determine the degree of robustness and the generalization ability of the suggested models, the dataset was separated into training and testing subsets using various splitting ratios. The purpose of using heterogeneous split ratios was to question the effect of the training data percentage on model behavior, in particular, the underlying class imbalance of the fraud detection problem. Three train test partitions were used in this research namely 70:30, 80: 20 and 90:10. All setups were made proportionate in terms of legitimate and fraudulent transaction in both training and testing set. The

RF algorithm acted as a control model throughout all the partitions to maintain uniform methodology and to monitor the effect of performance change with respect to the presence of data.

Table 1 Experimental Scores of Each Split of Dataset.

Ratio of Train and Test	**70:30**	**80:20**	**90:10**
Accuracy	0.9124	**0.9128**	0.9123

Table 1 shows that an 80:20 data partition had the highest accuracy (0.9128), which means that there is an optimal balance between an adequate training data and a representative test data. These results suggest that beyond a certain point, such as the one of a 90:10 division, the attained results are limited to small gains or even insignificant ones in the model performance. On the other hand, the drop in accuracy with a lower training ratio of 70:30 is small and this can be explained by inadequate exposure to diverse transaction patterns during this learning period. Experiments were performed and done as to the same preprocessing requirements which included scaling of features with MinMaxScaler and the selection of features using LASSO regression. To ensure reliability, this repeated training and evaluation of each model were done, and average estimates of the accuracy were then taken. The accuracy values in the different data splits have a small distribution, which highlights the stability and consistency of the Random Forest model in identifying the fraud and dubious transactions produced in the e-wallet dataset. The chosen ratio of 80: 20 in training and testing was then used in other experiments in which other machine-learning, deep-learning and hybrid modelling methods were used as it has better trade-off in terms of predictive ability and computational cost.

3.5. Dataset Suitability

Synthetic Mobile Money Transaction Dataset (PaySim) has been selected due to its scalability, the realistic aspect and giving transaction-level granularity that is essential in building fraud detection models within mobile payment systems. PaySim allows testing elegant structures, including hybrid convolutional neural networkrecurrent neural network (CNNRNN) systems, without invasion of user privacy and regulatory compliance, by statistically repeating realistic financial behaviors. In addition, the heterogeneity of categories of transactions in the dataset facilitates the generalization of the results in different situations of usage of e-wallets.

3.6. Machine Learning Model

The first application of machine learning (ML) algorithms was to set the performance metrics at the baseline when it comes to e-wallet fraud detection. These models have been picked because they are interpretable, computationally efficient and have proven to be effective in classification tasks. There were 5 supervised learning algorithms commonly used, trained, tuned, and tested, including, K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest and XGBoost. All the models are trained on an 80:20 training: test split, with the hyperparameters being optimized to obtain the best performance.

3.6.1. Random Forest

Random Forest algorithm is a collection of several decision trees, which uses bootstrap aggregation (bagging) to enhance the prediction ability and overcome overfitting. All constituent trees are developed on randomly subsampled data and a random set of features. The majority voting among the ensemble gives final predictions. The Random Forest was selected in the current research as a strong baseline classifier due to its resistance to noisy variables and the ability to predict nonlinearities within the dataset.

3.6.2. K-Nearest Neighbors

KNN algorithm uses the majority label of a sample in the feature space as the class label of the sample. This non-parametric approach is beneficial when dealing with data sets that have complicated and nonlinear boundaries of the classes. The proximity measure was the Euclidean distance and the best value of k was arrived at through experimentation.

3.6.3. Decision Tree

Decision Tree model is a hierarchical model which recursively divides the dataset based on features in terms of thresholds with the main objective of minimising impurity. In single use, it is prone to overfitting though it can be interpreted easily.

3.6.4. Logistic Regression

The Logistic Regression was applied as a statistical baseline to binary classification, between fraudulent and legitimate transactions. The model estimates the likelihood of fraud occurrence through a sigmoid activation function that is used to the linear combination of predictor variables*.*

3.6.5. Extreme Gradient Boosting

XGBoost is a gradient-boosting model, which make decision-trees in series, where each successive tree is designed to rectify the amount of its predecessor. Known to be highly efficient in computing, as well as in making predictions, XGBoost includes regularization terms to prevent overfitting.

3.7. Ensemble Model

In order to increase the level of predictive power as well as reduce variance, an ensemble model involving a combination of Random Forest and XGBoost was developed. Ensemble learning takes advantage of the complementary strengths of multiple models and thus provides a predictive performance that is better than what could be provided by individual classifiers.

3.7.1. Random Forest + XGBoost Ensemble

The ensemble technique is a combination of the probabilistic predictions of both the XGBoost and the Random Forest by using a weighted averaging model. This approach is useful in capturing heterogeneous boundaries in decisions made by every constituent model - the Random Forest endows the robustness to overfitting, and XGBoost provides high accuracy through gradient-boosted optimization. The ensemble thus achieved a top machine-learning accuracy of 0.91995, and this therefore validates that the hybridization of tree-based learners enhances fraud detection effectiveness. Accordingly, the model forms a powerful baseline on which future developments into deep-learning systems can be based.

3.8. Deep Learning Model

The transactional data were subjected to the deep machine learning (DL) techniques to automatically derive complex hierarchical features. In contrast to traditional machine-learning systems, the DL models have the capability of learning nonlinear relationships of high-dimensionality without requiring heavy manual feature engineering. We have used three models, the artificial neural network, the convolutional neural network and the recurrent neural network.

3.8.1. Alternative neural Network

The ANN model consisted of a series of completely interconnected (dense) layers using rectified linear unit (ReLU) activation along with a binary classification output layer using the sigmoid activation. In order to counter over fitting, dropout layers were added. Training was continued on the Adam optimizer with the loss criterion being binary cross-entropy. The final accuracy of 0.9221 is a testament to the ability of the model to identify some complicated correlations between transactional attributes.

3.8.2. Convolutional Neural Network

Although convolutional neural networks (CNNs) are traditionally used to work with spatial data, they can be used to work with structured data by viewing feature vectors as one-dimensional grids in practice. The CNN model used various convolutional and pooling layers, which allowed identifying local patterns between transaction features. The model achieved the maximum overall accuracy of 0.9224, thus demonstrating the ability of CNNs to identify local interaction of features that are relevant to fraudulent behavior.

3.8.3. Recurrent Neural Network

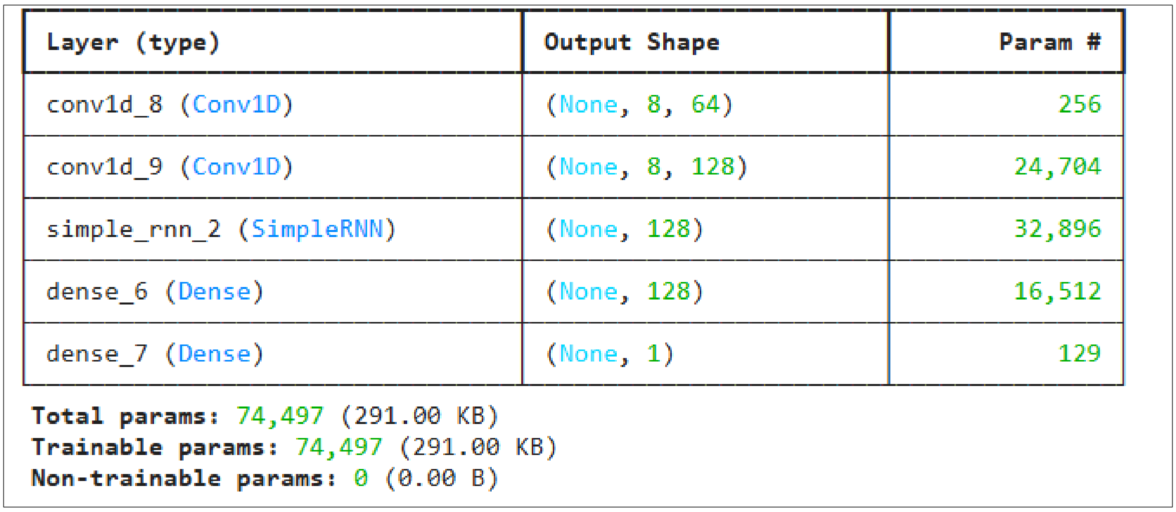
RNNs are particularly suitable with sequential or time-dependent data, whereby past history of transactions is used to predict the future. The RNN model took advantage of the use of recurrent layers to develop temporal relationships throughout the transaction sequence. Having an accuracy of 0.9178, RNNs were able to predict the sequential dynamics, but with slightly worse performance when compared to the CNN, which is due to the limited temporal context in the dataset.

3.9. Hybrid Learning Model

In order to use both spatial and sequence feature extraction, a hybrid CNN-RNN was built. This architecture is an integration of the feature-learning capability of convolutional neural networks and the temporal memory ability of recurrent neural networks, which create a more holistic framework of fraud detection.

3.9.1. CNN-RNN Hybrid

First, the hybrid model uses convolutional layers which isolate local spatial dependencies and learn high level patterns on the space of transaction features. The feature maps thus obtained are then inputted into an RNN layer that learns the sequential dependencies between consecutive transactions. The resulting output is taken through dense layers to give binary classification. The combined strategy enables the model to consider proximate correlations and chronological order in data concerning transactions at the same time. The hybrid CNN-RNN obtained accuracy of 0.9222, which is much more similar to the standalone CNN but it shows more generalization and learning stability on repeated design. These findings affirm that when several deep learning structures are combined together, a good stability between accuracy and recall can be achieved and hence a stronger system of detection of fraud is realized.



Layer (type)	Output Shape	Param #
conv1d_8 (Conv1D)	(None, 8, 64)	256
conv1d_9 (Conv1D)	(None, 8, 128)	24,704
simple_rnn_2 (SimpleRNN)	(None, 128)	32,896
dense_6 (Dense)	(None, 128)	16,512
dense_7 (Dense)	(None, 1)	129

```
Total params: 74,497 (291.00 KB)
Trainable params: 74,497 (291.00 KB)
Non-trainable params: 0 (0.00 B)
```

Figure 3 Architecture of Hybrid (CNN, RNN) Model

The architecture (shown in Figure 3) has five main layers, two convolutional layers, a single simple recurrent layer and two fully connected dense layers. This structure makes the model learn local pattern of features first and then learn sequential relationships among the transactions, which would increase the accuracy of fraud classification. The full hybrid model has 74 497 trainable parameters and it uses about 291 KB of memory, which implies a small and computationally efficient structure that makes it usable in real-time applications. These are all parameters that can be trained, thus making the model to be able to fit the distribution of data completely throughout the training process.

4. Performance evaluation

As stated in Table 2, our research has been effective to evaluate a wide variety of machine-learning, deep-learning and ensemble, as well as hybrid models. The results show the average results are always better in all the models, and Accuracy is used as the main measure of comparison. Specifically, the deep-learning model that showed the most accuracy of 0.9224 was the Convolutional Neural Network. The next in line were the Hybrid CNN+RNN and the deep-learning Artificial Neural Network model with the highest accuracy of more than 0.92 each. Although the vast majority of models, in particular those with high performance exceeding basic machine-learning methods, still maintained high and consistent performance in terms of Precision, Recall, F1 -Score (usually 0.91 -0.92), the dominant discriminative ability of the CNN highlights why it is the most effective architecture in this classification task.

Table 2 Model Performance Summary

Type of Learning	Model	Accuracy	Precision	Recall	F1-Score
Machine Learning	RF	0.9128	0.90	0.91	0.91
	KNN	0.9025	0.90	0.90	0.90
	DT	0.8956	0.89	0.90	0.89
	LR	0.8591	0.94	0.86	0.88
	XGB	0.9196	0.91	0.92	0.91
Ensemble Learning	RF+XGB	0.9199	0.91	0.92	0.91
Deep Learning	ANN	0.9221	0.91	0.92	0.91
	CNN	0.9224	0.91	0.92	0.91
	RNN	0.9178	0.91	0.92	0.91
Hybrid Learning	CNN+RNN	0.9222	0.91	0.92	0.91

The convolutional neural network (CNN) model proved to be highly efficient in the case of classifying the most represented class as represented in Figure 4. The number of seventeen thousand four hundred and twenty-seven instances that were correctly detected as class 0 (True Negatives) was correctly classified, and only 355 instances were falsely classified, as class 1 (False Positives). In the case of minority class 1, the model accurately established 1,021 (True Positives). However, it did not identify 1,197 cases, which were falsely identified as class 0 (False Negatives).

The high imbalance between the true negatives and the other measures shows that there is a significant skew in the classes in the dataset. Nevertheless, the true negatives outnumbered the primitives, which proves the effectiveness of the model in identifying class 0. Nonetheless, the large number of false negatives indicate that it could use more optimization to make the model recognize all false positives of class 1.

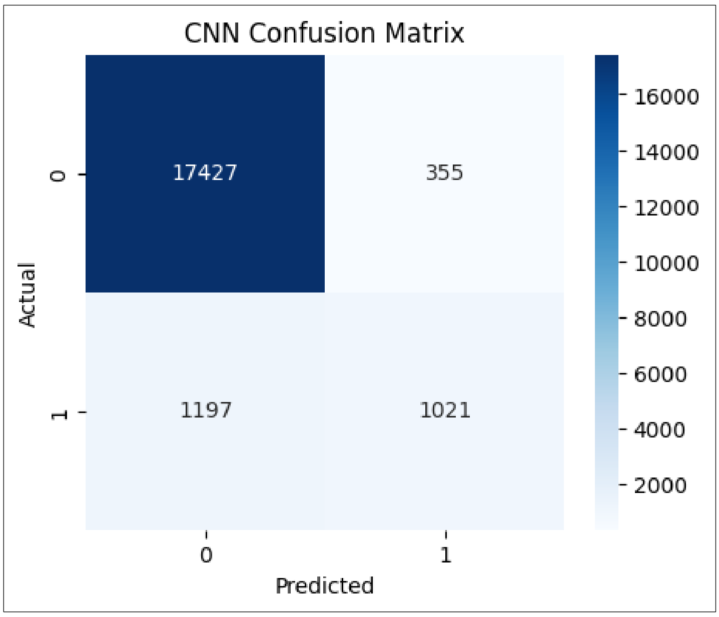


Figure 4 Confusion Matrix of CNN Model

Figure 5 revealed that both the train vs. validation accuracy and train vs. validation loss curves of the CNN model have high and stable learning dynamics in 100 training epochs. The Training Accuracy and Validation Accuracy values have an increasing pattern on the left panel up to about epoch 30 after which they level off and are closely related to one another, as the maximum value of 0.922 is reached. Notably, the validation accuracy is always quite close to the training accuracy, which implies that the model would generalise without severe overfitting. In line with this, the right panel shows both Training Loss and Validation Loss reducing drastically until the 30th epoch after which it levels off at approximately 0.152 to 0.155. The fact that the two sets of curves have been able to converge closely, especially the high final validation accuracy and low final validation loss are indications that the model has effectively converged and acquired robust features using the data.

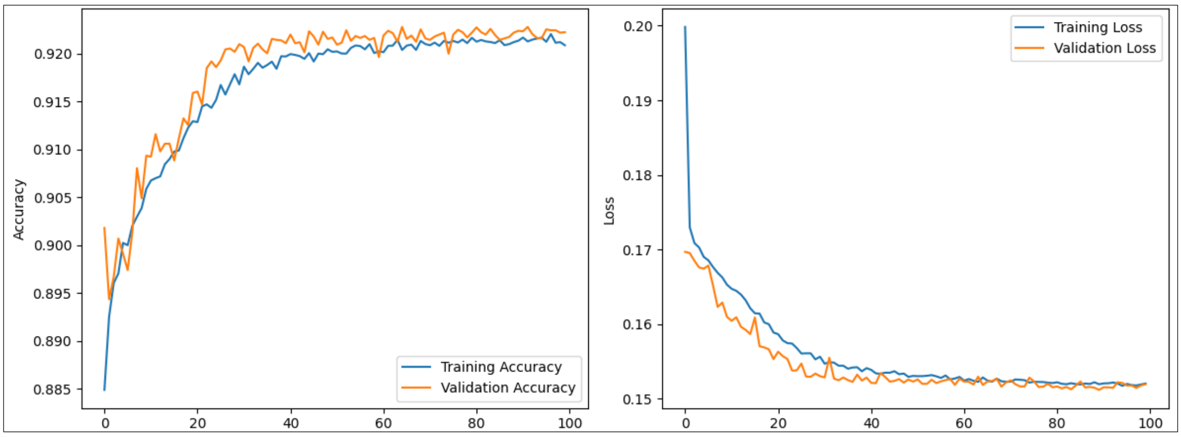


Figure 5 Accuracy vs Loss Curve for Training-validation of CNN Model

According to Figure 6 which shows the ROC curve, the convolutional neural network (CNN) has better discriminative ability. The curve which represents the True Positive Rate (Sensitivity) versus the False Positive rate (1-Specificity) rises rapidly at the beginning and reaches a true positive rate of 1.0 whilst confining the false positive rate to a lower value of 0.2. This trend means that the model is able to perform an extremely high recall that is, it identifies the majority of the positive cases with a very small penalty on the accuracy, thus minimizing the false alarms. The AUC of the curve is numerically obtained as 0.953. Since AUC is close to the theoretical limit of 1.0 and significantly higher than a random classifier (AUC = 0.5), this finding proves the strong performance of CNN and its ability to differentiate the two classes.

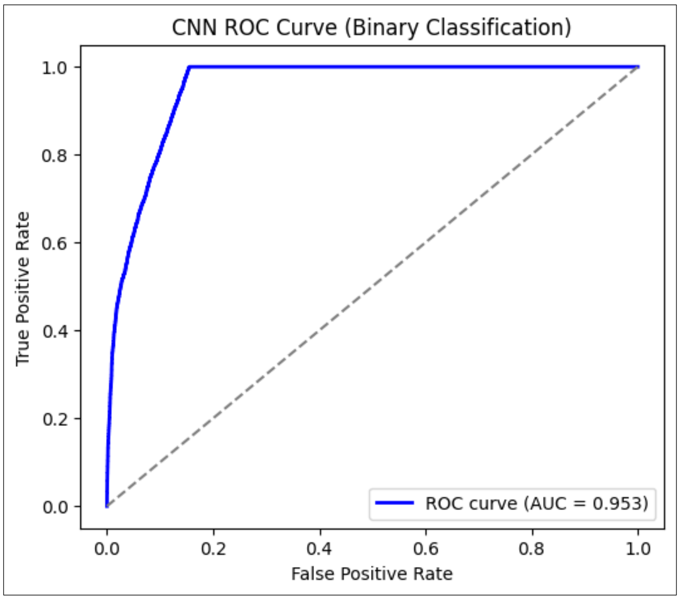


Figure 6 ROC Curve of CNN Model

5. Conclusion

In this research work, we are presenting a smart fraud detection system of e-wallet transactions that incorporates traditional methods of machine learning, ensemble learning, and deep learning techniques. Using the Synthetic Mobile Money Transaction Dataset (PaySim), we used a variety of predictive models to identify fraudulent activity in online payment systems. To ensure that the data is always consistent and the model is more powerful, the study employed widespread preprocessing processes like feature scaling, feature encoding, and feature selection through LASSO. Among the collection of conventional algorithms, XGBoost and Random Forest showed a higher predictive accuracy, and the combination of these classifiers further enhanced the detection. The capability to learn complex nonlinear correlation and transaction-behavioral pattern was demonstrated by the deep learning architectures, such as ANN, CNN, and RNN. The hybrid CNN-RNN system achieved a total accuracy of 92.22, which was higher than traditional machine-learning frameworks as well as separate deep-learning frameworks. These findings support the idea that hybrid architecture with the combination of spatial and time features extraction significantly improves the effectiveness of fraud detection in e-wallet contexts. The proven efficiency, scalability, and reliability of the model make it a feasible implementable model in the real-time in financial platforms. The presented methodology strengthens the security of transactions and enhances trust in digital transaction systems among consumers because of reducing false positives and improving the score of the detection process.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Jurgovsky J, Granitzer M, Ziegler K, Calabretto S, Portier P, He-Guelton L, Caelen O. Sequence classification for credit-card fraud detection. Expert Syst Appl. 2018;100:234–245.

[2] Carcillo F, Le Borgne Y-A, Caelen O, Bontempi G. Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection: Assessment and Visualization. arXiv:1804.07481. 2018.

[3] Dal Pozzolo A, Boracchi G, Caelen O, Alippi C, Bontempi G. Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy. IEEE Trans Neural Netw Learn Syst. 2018.

[4] Whitrow C, Hand D, Juszczak P, Weston D, Adams N. Transaction aggregation as a strategy for credit card fraud detection. Data Min Knowl Discov. 2009;18(1):30–55.

[5] Fiore U, De Santis A, Perla F, Zanetti P, Palmieri F. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Inf Sci. 2017.

[6] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: KDD’16 Proceedings. 2016.

[7] Carcillo F, Dal Pozzolo A, Le Borgne Y-A, Caelen O, Mazzer Y, Bontempi G. SCARFF: a Scalable Framework for Streaming Credit Card Fraud Detection with Spark. arXiv:1709.08920. 2017.

[8] Nguyen VB, et al. The Importance of Future Information in Credit Card Fraud Detection. arXiv. 2022.

[9] Zhao C, et al. Advancing financial fraud detection: Self-attention and related deep learning advances. Decis Support Syst. 2024.

[10] Sha Q, Tang T, Du X, Liu J, Wang Y, Sheng Y. Detecting Credit Card Fraud via Heterogeneous Graph Neural Networks with Graph Attention. arXiv:2504.08183. 2025.

[11] Chen Y, et al. Deep Learning in Financial Fraud Detection: systematic review (2019–2024). 2025.

[12] PaySim: PaySim synthetic mobile money transaction simulator / dataset (Kaggle).