Ensemble Learning Approach to Fraud Detection in Cryptocurrency Blockchain

Alowolodu Olufunso Dayo
Department of Cybersecurity
Federal University of Technology
Akure, Nigeria

ABSTRACT
Blockchain, an emerging and very important technology in the financial industry is facing many challenges especially security wise. The decentralized nature and characteristics of the blockchain makes it more difficult for conventional intrusion detection and prevention systems to identify and prevent fraudulent activities in real-time. This has posed serious challenges for fraud detection systems, thereby contributing to the wider attempts being made to ensure secure blockchain environments and build trust in cryptocurrency markets. This research hereby proposes an ensemble model approach to detect fraudulent cryptocurrency transaction. The proposed model will combine two deep learning algorithms namely, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM). The ensemble model consistently demonstrated high precision and at the same time ensured that the transactions that were labelled fraudulent were indeed captured as true, while sustaining high recall to identify most of the fraudulent activities. This work has shown that ensemble learning can generate a more robust and accurate fraud detection system rather than the conventional or single models and this makes the model more relevant in situations with highly imbalanced datasets like cryptocurrency transactions like blockchain.

General Terms
Blockchain Security

Keywords
Blockchain, Cryptocurrency, Deep Learning, Fraud, Security.

1. INTRODUCTION
The emergence of cryptocurrency has created a significant change in the financial world by enabling a secured, decentralized, and borderless transactions. This phenomenon has been classified into various categories like Bitcoin and Ethereum which has enabled users to transact without fear of privacy invasion, identity disclosure, and has also given room for autonomy and financial inclusion of the end users (Conti et al., 2018). Despite all the added benefits of this phenomenon, it was pillaged by series of challenges especially in the area of security and fraud detection. One of the most prevalent areas of application of this is the Blockchain and this has made cryptocurrencies a level ground for unlawful and illegal activities like money laundering, Ponzi schemes, and phishing domain attacks (Foley et al., 2019). Due to the decentralized nature and characteristics of the blockchain, it is more difficult for conventional intrusion detection and prevention systems to identify and prevent fraudulent activities in real-time, and this is made more compounded by the pseudonymity of users (Chen et al., 2020). This has necessitated the introduction of more sophisticated fraud and intrusion detection systems in order to uphold the integrity and trustworthiness of transactions on the block (Hussaini et al., 2022).

This research hereby proposes an ensemble model approach to detect fraudulent cryptocurrency transaction. This model will combine two deep learning algorithms namely: Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks as they are most suitable for analyzing sequential time-series data, which is important for detection of temporal patterns in cryptocurrency transactions (Taher, 2024). The proposed model will ensure scalability, adaptability, and fraud detection accuracy of cryptocurrency transactions, hereby contributing to the wider attempts being made to ensure secured blockchain environments and build trust in cryptocurrency markets (Kim et al., 2023).

2. RELATED WORKS

2.1 Blockchain and Cryptocurrency Transactions
Cryptocurrency has many areas of application with Blockchain technology being one major area. This Blockchain technology is a decentralized digital and distributed ledger system that records transactions across multiple computers, hereby ensuring security, transparency and immutability of records (Nakamoto, 2008, Tapscott & Tapscott, 2016). Cryptocurrency transactions encompass the transfer of digital assets by means of cryptographic algorithms that contains both public and private keys. The public key is meant for receiving of funds while the private key is meant for authorizing transactions. In spite of the security attribute of the blockchain technology, the system is not invulnerable to fraudulent activities, like phishing, Ponzi schemes, double-spending, exchange hacks, and ransomware (Bertoni, 2017; Nakamoto, 2010; Kharif, 2014; Goodin, 2017). To mitigate these challenges, there is need for a more sophisticated fraud detection and prevention mechanisms to ensure the integrity, confidentiality of the blockchain environments.

2.2 Fraud Detection in Blockchain
The Blockchain technology is fundamentally a decentralized digital ledger that ensures that data once recorded, cannot be altered or tampered with without the consensus of the network. Coupled with the pseudonymous nature of the transactions, this has posed serious challenges for fraud detection systems.

Conventional fraud detection techniques like the rule-based systems, statistical methods, and machine learning models still struggle with the evolving and intricate fraud patterns that characterized the blockchain environments (Chandola et al., 2009; Ngai et al., 2011; Han et al., 2011). The work of Quadir et al. (2023) that proposed a Novel Approach to Detecting Fraud in Ethereum Transactions Using Stacking by the combination of various classifiers to improve fraud detection performance by leveraging on different strengths of the combined models. Improved detection rates and reduced false positives were recorded. Even though the approach does not require any substantial computational resources, the model

developed may not be easily scalable. Pahuja and Kamal (2023) proposed an Ensemble Learning-Based Ethereum Fraud Detection Using CRISP-DM Framework to detect fraudulent activities in Ethereum transactions. The model could not perform as expected due to the heavy reliance on the quality and relevance of the input features.

Airlangga (2024) proposed an Anomaly Detection in Blockchain Transactions within the Open Metaverse Using Unsupervised Learning within decentralized metaverse environments. The paper highlighted the ability of unsupervised methods to identify unusual patterns without labelled data hereby addressing fraud detection in a new and evolving context, though it acknowledges the limitations of unsupervised learning in handling diverse and complex transaction patterns. These showed the efficiency of Machine learning models, particularly ensemble techniques in improving fraud detection accuracy within the context of cryptocurrencies. Additionally, the need for real-time detection is a driving factor. Conventional fraud detection systems often experience delays and high false-positive rates, which hinder their ability to prevent financial losses in cryptocurrency transactions (Khan and Alshahrani, 2023). The stacking ensemble method combining RNN and LSTM models will improve real-time detection accuracy, helping to maintain user trust and security within blockchain ecosystems (Nguyen & Nguyen, 2023). Despite these efforts and many more, challenges like scalability, data privacy, and model adaptability still continue to serve as obstacles.

2.3 Machine Learning Approaches

Recent literatures have explored the use of machine learning models for blockchain fraud detection. Saxena et al. (2024) proposed an ensemble learning technique for blockchain transaction deanonymization, while the work of Quadir et al. (2023) used stacking classifiers to improve fraud detection in Ethereum transactions. Several other approaches, like the work of Pahuja and Kamal (2023) and Airlangga (2024), have also proved the efficiency of ensemble methods and unsupervised learning in blockchain fraud detection. Despite the success of these methods, scalability and ability to detect new and evolving fraud patterns in real-time still serve as challenges.

3. METHODOLOGY

3.1 Data Collection and Preprocessing

The dataset employed for this work was sourced from Kaggle and consists of 9,841 rows, with 51 features, each labeled as either fraudulent or legitimate activities. The Data preprocessing steps are as followed:

i. **Missing Data**: Missing values were assigned using the mean for numerical features and the mode for categorical variables using the mean imputation as shown in equation 1

$$X_{imputed} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad eqn\ 1$$

Where:
$X_{imputed}$ is the value used to replace the missing data.
X_i are the observed values of the feature.
n is the number of non-missing observations.

ii. **Encoding Categorical Variables**: Categorical variables were one-hot encoded to transform them into a numerical format suitable for machine learning models.

iii. **Normalization**: Numerical features were normalized using the StandardScaler to ensure uniform scaling across all features by transforming them to have a mean of 0 and a standard deviation of 1 which is given as:

$$X_{scaled} = \frac{X_{max} - X_{min}}{X - X_{min}} \qquad eqn\ 2$$

Where:

X_{scaled} is the normalized value, X is the original feature value, $X_{min}\ and\ X_{max}$ are the minimum and maximum values of the feature, respectively.

The dataset was then split into training and testing sets, ensuring that the models were evaluated on unseen data.

3.2 Feature Engineering and Selection

Feature engineering was performed to enhance the model's predictive power. New features, such as transaction frequency and amount variability, were derived from the original dataset. Feature selection was conducted using a Random Forest model, which identified the most important features for fraud detection. These selected features were then standardized to ensure uniform scaling and to optimize model performance.

3.3 Synthetic Minority Over-sampling (SMOTE)

To address class imbalance, where fraudulent transactions were vastly outnumbered by legitimate ones, SMOTE was applied. This technique generates synthetic fraudulent transactions by interpolating between instances of the minority class, balancing the dataset and improving model learning.

3.4 Ensemble Model Design

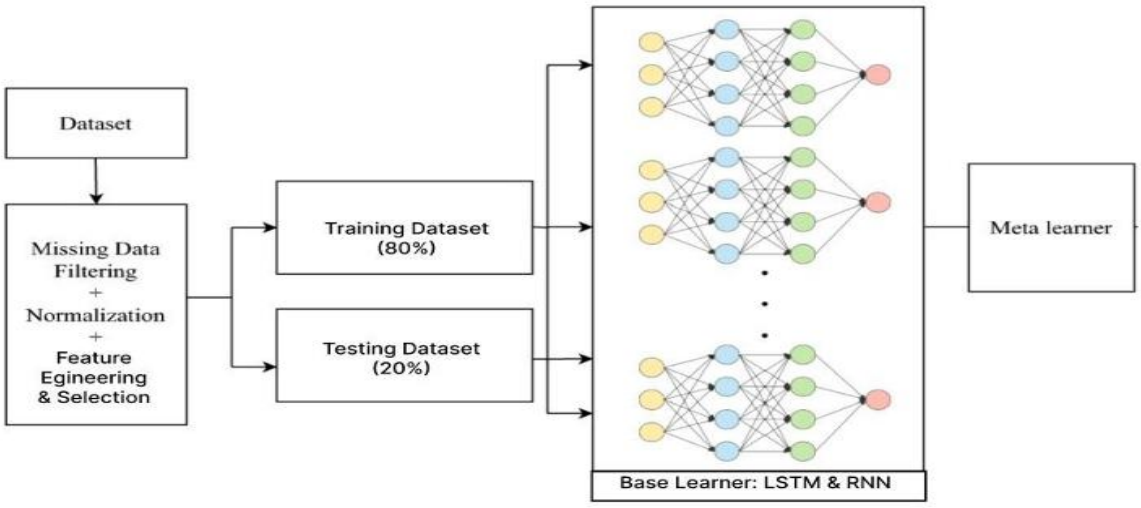

Figure 1: The Proposed Ensemble Architecture

The proposed system was designed using a stacking ensemble method that combines the predictions of RNN and LSTM models and leveraged on their complementary strengths:

i. **RNNs**: RNNs are designed to capture short-term dependencies in transactional data, making them effective for detecting immediate patterns indicative of fraud making them a strong choice for analysing cryptocurrency transactions, which occur in a time-series format. The RNN model is used to detect **short-term dependencies** in transactional data, focusing on recent transaction patterns that may indicate fraudulent behaviour which are represented by the:

a. **Hidden State Update:**

$$h_t = tanh\,(W_h h_{t-1} + U_h x_t + b_{h)} \quad eqn\,3$$

Equation 3 above Updates the hidden state h_t based on the previous hidden state h_{t-1} and current input x_t .

b. **Output Computation:**

$$y_t = \sigma\big(W_h h_t + b_y\big) \quad eqn\,4$$

Produces the output y_t from the hidden state h_t using a dense layer and activation function σ

c. **Loss Function:**

$$Loss = -\frac{1}{T}\sum_{t=1}^{T}[y_t\,log(y_t) + (1-y_t)\,log(1-y_t)] \quad eqn\,5$$

Measures the error between the true labels y_t and the predicted probabilities y_t ,typically using binary cross-entropy for classification. In essence, the RNN updates its hidden state over time and uses it to generate predictions, with training focused on minimizing the error between predicted and actual values.

ii. **LSTMs**: LSTM networks, designed to overcome the vanishing gradient problem, are particularly effective at capturing long-term dependencies, which are crucial for detecting complex fraud patterns in cryptocurrency transactions.

The predictions from both models were combined into a meta-model, such as logistic regression, which optimized the final fraud detection output for improved accuracy ability to handle diverse patterns through:

i. **Cell State Update:**

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t \quad eqn\,6$$

Updates the cell state C_t using the forget gate f_t, previous cell state C_{t-1}, input gate C_{t-1}, and candidate cell state C_t.

ii. **Hidden State Update:**

$$h_t = 0_t\,tanh \odot (C_t) \quad eqn\,7$$

iii. **Gate Computations:**

i. **Forget Gate:**

$$f_t = \sigma\big(W_f\,[h_{t-1}, x_t] + b_f\big) \quad eqn\,8$$

ii. **Input Gate:**

$$f_t = \sigma\big(W_f\,[h_{t-1}, x_t] + b_i\big) \quad eqn\,9$$

a. **Candidate Cell State:**

$$f_t = \sigma\big(W_f\,[h_{t-1}, x_t] + b_c\big) \quad eqn\,10$$

b. **Output Gate:**

$$f_t = \sigma\big(W_f\,[h_{t-1}, x_t] + b_c\big) \quad eqn\,11$$

ii. **Loss Function:**

$$Loss = -\frac{1}{T}\sum_{t=1}^{T}[y_t\,log(y_t) + (1-y_t)\,log(1-y_t)] \quad eqn\,12$$

4. RESULTS AND DISCUSSION

4.1 Evaluation Metrics

The performance of the models was evaluated using the following metrics:

i. Accuracy: The proportion of correctly classified transactions represented by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad eqn\ 13$$

ii. Precision: The proportion of true fraudulent transactions identified out of all transactions predicted as fraudulent. This is represented by:

$$Precision = TP/(TP + FP) \quad eqn\ 14$$

iii. Recall: The proportion of true fraudulent transactions identified out of all actual fraudulent transactions using equation 15

$$Recall = TP/(TP + FN) \quad eqn\ 15$$

iv. F1-Score: The harmonic mean of precision and recall, providing a balanced measure of performance. It is the harmonic mean of precision and recall, and a higher F1 score that indicates a good balance between the two and this is represented in equation 16 below

$$F1 - score = 2\ X\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad eqn\ 16$$

AUC-ROC: The area under the receiver operating characteristic curve, which measures the model's ability to discriminate between fraudulent and legitimate transactions is shown as

$$PR = FP/(FP + TN) \quad eqn\ 17$$

$$AUC - ROC = \int_0^1 TPR(FPR)d(FPR) \quad eqn\ 18$$

4.2 Model Performance

Class Distribution After Smote

After applying Synthetic Minority Over-sampling Technique (SMOTE) on the dataset, the distribution of fraudulent and non-fraudulent transactions in the dataset becomes more balanced by generating synthetic samples of the minority class (fraudulent transactions), ensuring that both classes have roughly equal representation.

Before SMOTE, the dataset was heavily skewed towards non-fraudulent transactions, which made it challenging for the model to learn patterns associated with fraud. After SMOTE, the number of fraudulent transactions is artificially increased by interpolating between existing data points of the minority class.

4.3 Model Performance

Class Distribution After Smote

The Blockchain technology is fundamentally a decentralized After applying **SMOTE (Synthetic Minority Over-sampling Technique)**, the distribution of fraudulent and non-fraudulent transactions in the dataset becomes more balanced by generating synthetic samples of the minority class (fraudulent transactions), ensuring that both classes have roughly equal representation.

Before SMOTE, the dataset was heavily skewed towards non-fraudulent transactions, which made it challenging for the model to learn patterns associated with fraud. After SMOTE, the number of fraudulent transactions is artificially increased by interpolating between existing data points of the minority class.

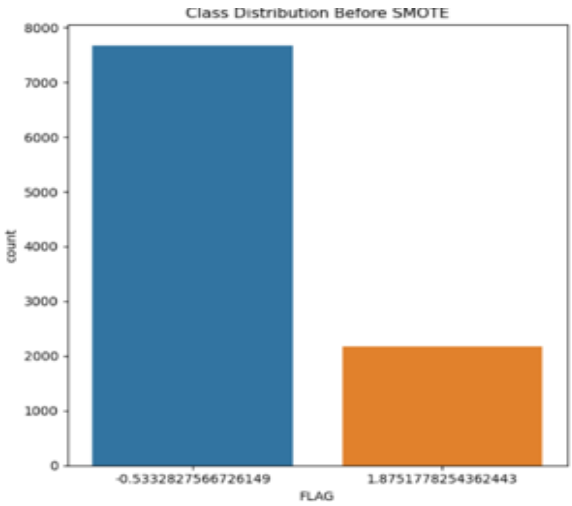

Figure 2a: Class Distribution Before SMOTE

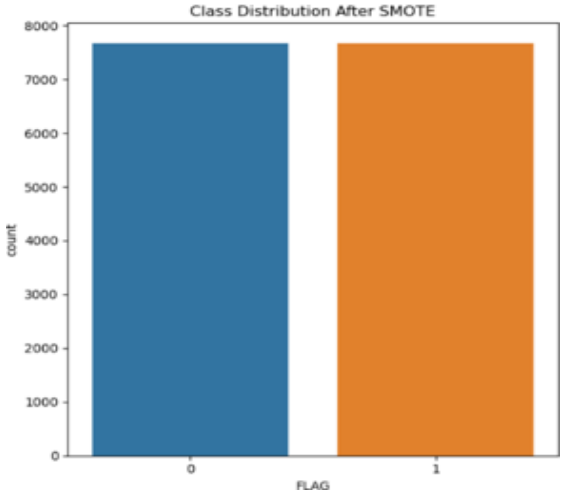

Figure 2b: Class Distribution After SMOTE

Figures 2a and 2b show the class distributions before and after the application of SMOTE on the datasets respectively. The count of the fraudulent dataset in Figure 2a was low while Figure 2b shows an increase in the count of the fraudulent dataset, which equals the count of the non-fraudulent dataset.

4.4 LSTM Model Evaluation

After training, the LSTM model was evaluated on the test set to assess its performance on unseen data The model achieved an accuracy of 97.65% and a precision of 94% on the test set. The binary cross-entropy loss was also reported. The trained model was saved for future use in detecting fraudulent transactions.

LSTM Classification Report:

	precision	recall	f1-score	support
Legitimate	0.97	0.96	0.98	7662
Fraudulent	0.93	0.97	0.98	7662
accuracy			0.97	15324
macro avg	0.98	0.98	0.98	15324
weighted avg	0.95	0.96	0.98	15324

Figure 3: LSTM Accuracy

The training process was visualized via the training and validation losses plot and training and validation accuracies. Plot. The plots help in understanding the model's performance over time and identifying any potential overfitting.

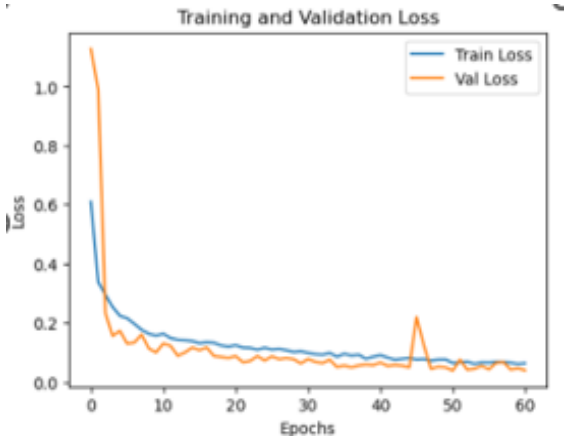

Figure 4a: LSTM Loss Plot

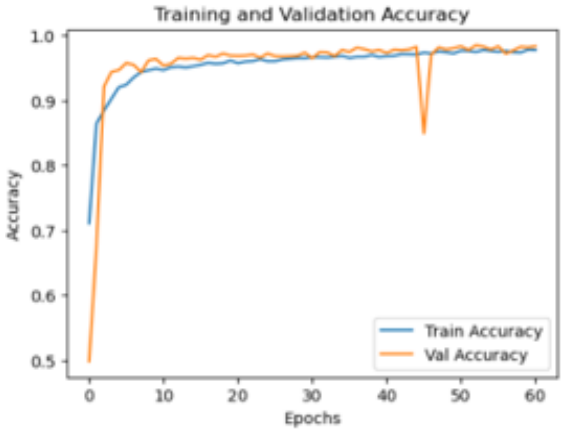

Figure 4b: LSTM Accuracy Plot

The training and the validation loss plot indicates change over the epochs. A decreasing validation loss indicates that the model is improving during training while the accuracy plot monitors the model's accuracy on both the training and validation sets, hereby giving a visual representation of how the model reacts to unseen data.

4.5 RNN Model Evaluation

The RNN model was used to access the test data in order to evaluate its performance on unseen data. The results showed an accuracy of 97% and a precision of 95% on the test set.

```
479/479 ━━━━━━━━━━━━━━━━━━━━ 7s 9ms/step
RNN Classification Report:
              precision    recall  f1-score   support

  Legitimate       0.91      0.98      0.97      7662
  Fraudulent       0.98      0.96      0.97      7662

    accuracy                           0.97     15324
   macro avg       0.97      0.97      0.97     15324
weighted avg       0.94      0.97      0.97     15324
```

Figure 5: RNN Accuracy

To visualize the training process, the training and validation losses as well as the accuracies were plotted in Figures 6a and 6b to further provide insights into the model's convergence and help in detecting potential overfitting.

Figure 6a: RNN Loss Plot

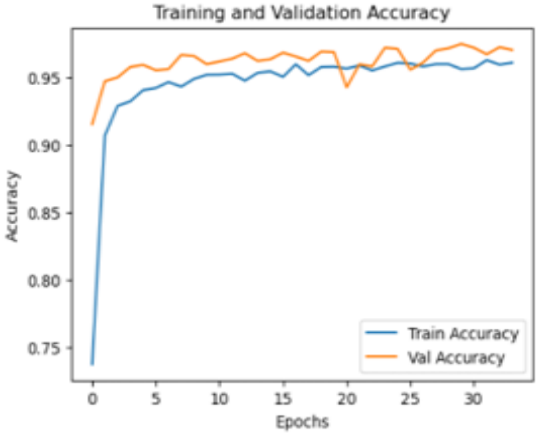

Figure 6b: RNN Accuracy Plot

4.6 Ensemble Model Evaluation

To evaluate the performance of the ensemble model, standard metrics like **accuracy**, **precision**, **recall**, **F1-score**, and **ROC AUC** were employed. Moreso, the **confusion matrix** was also generated to determine the model's performance.

```
Accuracy: 0.9785
Precision: 0.9804
Recall: 0.9765
F1 Score: 0.9785
ROC AUC: 0.9987

Confusion Matrix:
[[1500   30]
 [  36 1499]]

Classification Report For RNN & LSTM combined as Ensemble Stack:
              precision    recall  f1-score   support

           0       0.98      0.98      0.98      1530
           1       0.98      0.98      0.98      1535

    accuracy                           0.98      3065
   macro avg       0.98      0.98      0.98      3065
weighted avg       0.98      0.98      0.98      3065
```

Figure 7: Ensemble Model Metrics

From Figure 7 above, an accuracy of 0.98, precision of 0.98, f1 score of 0.978, ROC and AUC of 0.99, and recall of 0.976 were recorded from the training of the meta learner.

To further access the performance of the model, the ROC curve and a heatmap of the confusion matrix were generated. The **ROC curve** was to evaluate the model's ability to discriminate between fraudulent and legitimate classes while the area under the curve (AUC) provided a single measure of this performance. The ROC curve was plotted to visualize the disparity between the false positive and the true positive rate and also differentiate between the fraudulent and legitimate transactions.

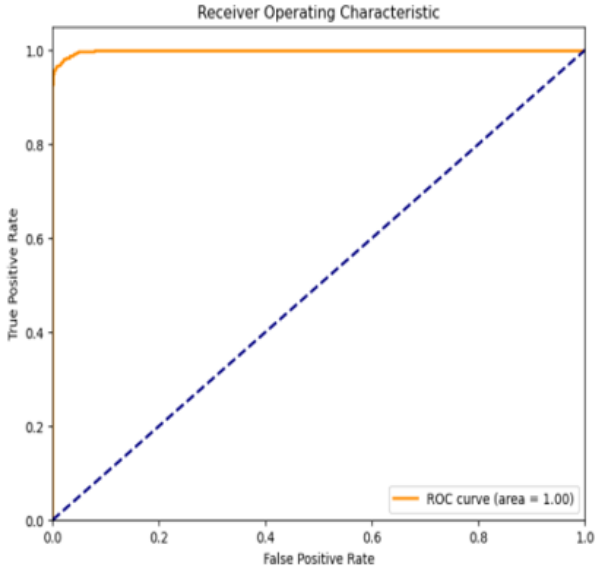

Figure 8: The Receiver Operating Characteristic (ROC) curve

In figure 8, the ROC curve area is 1.0 which shows a very good prediction performance of the model. **A heatmap of the confusion matrix** graph as represented in figure 8 represented the number of true positives, true negatives, false positives, and false negatives to give a clearer insight to the model's performance. The heat map was used to visualize the confusion matrix using the color intensity to represents the frequency of correct and incorrect predictions.

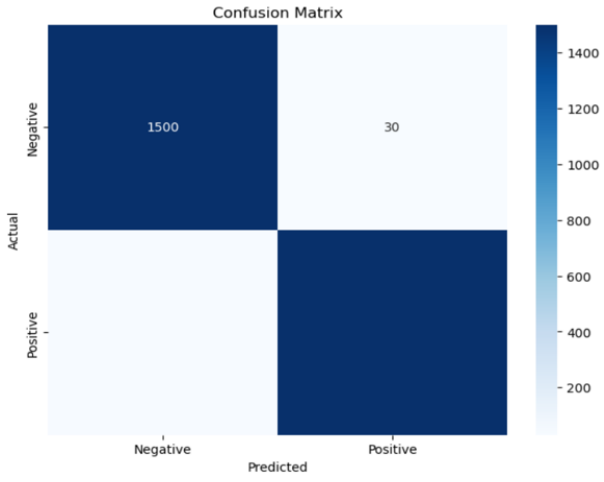

Figure 9: Heatmap of the Confusion Matrix

In figure 9, the figure between the actual and predicted values are balanced due to the increase in the negative part of it having a high frequency. Summarily, the

i. **LSTM Model** achieved an accuracy of 97.65% and precision of 94%, demonstrating its effectiveness in capturing long-term fraud patterns while the
ii. **RNN Model**: Achieved an accuracy of 97% and precision of 95%, focusing on short-term patterns indicative of fraud. However, the **Ensemble Model** which combines both RNN and LSTM outputs, achieved an accuracy of 98%, precision of 98%, recall of 97.6%, F1-score of 97.8%, and AUC-ROC of 99.87%.

This demonstrates the effectiveness of the ensemble approach in capturing both short-term and long-term fraud patterns, enhancing overall detection accuracy in a cryptocurrency blockchain .

5. CONCLUSION

The ensemble model that combines the predictions of both the LSTM and RNN models has proven to be an effective model algorithm for the detection of fraudulent transactions on the blockchain. The RNN model was able to identify short-term dependencies in the transactional data, while the LSTM model was able to successfully identify the long-term patterns that may emerge over time. The model presented an improved overall accuracy and successfully reduced the classification of false positives and false negatives.

The ensemble model consistently demonstrated high precision and at the same time ensured that the transactions that were labelled fraudulent were indeed captured as true, while sustaining high recall to identify most of the fraudulent activities.

Finally, this work has shown that ensemble learning can ge4nerate a more robust and accurate fraud detection system rather than the conventional or single models and this makes the model more relevant in situations with highly imbalanced datasets like cryptocurrency transaction.

6. REFERENCES

[1] Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[2] Bartoni, A. (2017). *Ethereum phishing attack results in $70 million theft. Forbes*. Retrieved from https://www.forbes.com/sites/antonyleather/2017/06/20/ethereum-phishing-attack-results-in-70-million-theft/Buterin, V. (2013). Ethereum White Paper. Retrieved from https://ethereum.org/en/whitepaper/

[3] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58. https://doi.org/10.1145/1541880.1541882

[4] Chen, Y., Li, N., Zhang, J., & Shao, L. (2020). Survey of fraud detection research on blockchain. Proceedings of the 2020 International Conference on Big Data and Artificial Intelligence (BDAI).

[5] Conti, M., Kumar, S., Lal, C., & Ruj, S. (2018). A survey on security and privacy issues of bitcoin. IEEE Communications Surveys & Tutorials, 20(4), 3416-3452.

[6] Foley, S., Karlsen, J. R., & Putniņš, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? The Review of Financial Studies, 32(5), 1798-1853.

[7] Goodin, D. (2017). *WannaCry ransomware attack hits 150 countries. Ars Technica*. Retrieved from

https://arstechnica.com/information-technology/2017/05/wannacry-ransomware-attack-hits-150-countries/

[8] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers. ISBN: 978-0123814791

[9] Hussain, Y., Dai, T., Ti, Y., Huang, M., Chiang, T., & Liu, L. (2022). Feature Engineering and Resampling Strategies for Fund Transfer Fraud With Limited Transaction Data and a Time-Inhomogeneous Modi Operandi. *IEEE Access*,10, 86101-86116. https://doi.org/10.1109/ACCESS.2022.3199425.

[10] Kharif, O. (2014). *Mt. Gox says it lost 850,000 Bitcoins in hack*. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2014-02-28/mt-gox-says-it-lost-850-000-bitcoins-in-hack

[11] Kim, J., Lee, S., Kim, Y., Ahn, S., & Cho, S. (2023). Graph learning-based blockchain phishing account detection with a heterogeneous transaction graph. Sensors, 23(1), 463. https://doi.org/10.3390/s23010463

[12] Kumar, A., & Sharma, I. (2023). Preserving Security of Crypto Transactions with Machine Learning Methodologies. 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), 129-134. https://doi.org/10.1109/ICSCSS57650.2023.10169192.

[13] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

[14] Nakamoto, S. (2010). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Retrieved from https://bitcoin.org/bitcoin.pdf

[15] Ngai, E. W. T., Xiu, L., & Chau, K. Y. (2011). Application of data mining techniques in fraud detection: A classification framework. Decision Support Systems, 50(3), 559-570. https://doi.org/10.1016/j.dss.2010.08.005

[16] Pahuja, L., & Kamal, A. (2023). Enlefd-dm: ensemble learning based ethereum fraud detection using crisp-dm framework. Expert Systems, 40(9). https://doi.org/10.1111/exsy.13379

[17] Quadir, A., Narayanan, S., Sabireen, H., Sivaraman, A., & Tee, K. (2023). A novel approach to detect fraud in ethereum transactions using stacking. Expert Systems, 40(7). https://doi.org/10.1111/exsy.13255

[18] Saxena, R., Arora, D., Nagar, V., & Chaurasia, B. K. (2024). Blockchain transaction deanonymization using ensemble learning. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-19233-5

[19] Taher, S. (2024). Advanced fraud detection in blockchain transactions: an ensemble learning and explainable AI approach. Engineering Technology & Applied Science Research, 14(1), 12822-12830. https://doi.org/10.48084/etasr.664.