Hybrid Autoencoder–LSTM Deep Learning Model for Detecting Anomalous Behavior in Metaverse Financial Transactions

Amaleswari Rajulapati[1,*], Sridevi Veerappan[2]

[1,2]Department of EEE, AMET Deemed to be University, ECR, Kanathur, Chennai, India

ABSTRACT

The expansion of the metaverse has created new possibilities for digital financial interaction while simultaneously increasing exposure to fraudulent and anomalous activities. This study introduces a Hybrid Autoencoder–LSTM Deep Learning Model designed to detect irregular transaction behaviors within metaverse financial systems. The model integrates an Autoencoder for nonlinear feature extraction and an LSTM network for sequential dependency learning, enabling comprehensive detection of behavioral and transactional anomalies. Using a dataset of 78,600 records encompassing demographic, behavioral, and financial attributes, the model achieved an overall accuracy of 79.76 percent and an Area Under the Curve (AUC) value of 0.742, demonstrating moderate capability in distinguishing legitimate from fraudulent transactions. The convergence of training and validation losses confirmed robust generalization without overfitting, while classification metrics revealed strong precision and recall for legitimate transactions but reduced sensitivity to rare anomalies due to class imbalance. These results validate the effectiveness of the proposed hybrid model as a foundation for intelligent fraud detection in virtual financial environments and highlight the importance of integrating deep learning and explainable AI approaches to strengthen security, transparency, and trust within metaverse economies.

Keywords Metaverse, Fraud Detection, Deep Learning, Autoencoder–LSTM, Anomaly Detection

INTRODUCTION

The metaverse has emerged as a transformative digital environment that integrates virtual reality, blockchain, and artificial intelligence to create immersive, interactive, and decentralized ecosystems. Within this space, users engage in a variety of activities such as trading virtual assets, purchasing digital goods, attending events, and participating in financial exchanges that mirror real-world economic systems. The increasing adoption of virtual currencies and non-fungible tokens (NFTs) has expanded the scope of financial transactions in the metaverse [1]. However, the rapid growth of these activities has also introduced new vulnerabilities, particularly in relation to fraudulent transactions and financial manipulation. As transactions in the metaverse often involve high-frequency exchanges across distributed networks, ensuring their integrity and authenticity has become a significant challenge for developers, regulators, and users [2].

The growing complexity of metaverse financial ecosystems has made traditional fraud detection techniques insufficient. Rule-based systems and manual transaction monitoring, which rely on predefined thresholds

Submitted: 12 May 2026
Accepted: 25 July 2026
Published: 17 November 2026

Corresponding author
Amaleswari Rajulapati,
rajulapatiamaleswari2025@gmail.com

Additional Information and Declarations can be found on page 312

DOI: 10.47738/jdmdc.v3i4.103

Copyright
2026 Rajulapati and Veerappan

Distributed under
Creative Commons CC-BY 4.0

How to cite this article: A. Rajulapati and S. Veerappan, "Hybrid Autoencoder–LSTM Deep Learning Model for Detecting Anomalous Behavior in Metaverse Financial Transactions," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 4, pp. 299-314, 2026.

or human intervention, struggle to identify sophisticated and evolving patterns of fraud [3]. These conventional approaches are limited in adaptability and cannot effectively handle the nonlinear, temporal, and multidimensional nature of metaverse data. As a result, anomalies or fraudulent behaviors that emerge gradually or occur in complex behavioral sequences often go undetected. Furthermore, the lack of centralized control and the pseudonymous nature of blockchain-based systems amplify these challenges, making it difficult to trace fraudulent actors. Consequently, the development of intelligent, data-driven, and adaptive fraud detection frameworks is essential to safeguard financial trust within the metaverse [4].

Recent advancements in artificial intelligence and deep learning have significantly enhanced the capabilities of fraud detection systems across financial and digital domains. State-of-the-art models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and Autoencoders have been successfully applied in various anomaly detection contexts [5]. LSTM models are particularly effective in capturing temporal dependencies in sequential data, making them suitable for detecting evolving behavioral patterns in transaction sequences. Autoencoders, on the other hand, excel at learning nonlinear feature representations and identifying deviations from normal data distributions. Existing research has shown promising results in the use of hybrid deep learning architectures for fraud detection in banking systems, credit card transactions, and cryptocurrency platforms. However, studies applying similar techniques within metaverse environments remain scarce. Unlike traditional financial datasets, metaverse transactions are characterized by a combination of behavioral, demographic, spatial, and temporal factors that require more complex modeling strategies.

Although prior research has demonstrated progress in AI-based fraud detection, several critical research gaps remain unaddressed. Many existing studies rely on static or structured datasets, which fail to capture the dynamic and immersive nature of metaverse economies. Models that focus solely on numerical transaction data often neglect behavioral or contextual indicators, such as session duration, login frequency, or purchase patterns, that are crucial for understanding user intent and detecting sophisticated fraud. Additionally, few models effectively integrate nonlinear and temporal features within a single architecture. As a result, current systems tend to exhibit high accuracy in detecting normal transactions but limited sensitivity in identifying rare and subtle anomalies. This imbalance highlights the need for hybrid deep learning models capable of simultaneously learning complex feature hierarchies and temporal dependencies to improve anomaly detection performance in metaverse transaction environments.

To address these limitations, this study proposes a Hybrid Autoencoder–LSTM Deep Learning Model for detecting anomalous behaviors in

metaverse financial transactions. The Autoencoder component captures nonlinear feature interactions, while the LSTM layers model temporal dependencies in user activity sequences. This combination allows the proposed model to reconstruct normal transaction patterns and identify deviations that may represent fraudulent behaviors. The study utilizes a dataset containing behavioral, demographic, and financial variables to train and evaluate the model's effectiveness. The objectives of this research are to assess the performance of the proposed hybrid model, compare its results with conventional approaches, and demonstrate its potential to enhance security and trust within virtual financial systems. The findings are expected to contribute to the development of intelligent, explainable, and adaptive AI-based fraud detection frameworks that support transparency, accountability, and resilience in metaverse financial ecosystems.

Literature Review and Related Works

The increasing volume and complexity of digital transactions in the metaverse have led to growing concerns regarding fraud and financial anomalies. Traditional rule-based systems are limited in their ability to detect sophisticated fraudulent patterns, especially when dealing with high-dimensional, dynamic, and decentralized data structures [6]. Consequently, artificial intelligence and deep learning approaches have emerged as powerful alternatives, offering the capacity to model complex behaviors and uncover hidden relationships within transactional data [7]. Among these, deep neural networks and hybrid learning models have shown the ability to handle the nonlinearity and temporal dependencies of financial transaction sequences [8].

Recent research has applied various deep learning techniques to anomaly and fraud detection tasks. Autoencoders and Long Short-Term Memory (LSTM) networks are two of the most widely utilized architectures in this context. Autoencoders are effective for unsupervised anomaly detection, as they reconstruct normal data patterns and identify deviations based on reconstruction errors [9]. LSTM networks, designed to model sequential dependencies, are particularly suitable for transaction data that exhibit time-series characteristics [10]. Comparative analyses have shown that hybrid models that combine Autoencoder and LSTM components can achieve higher detection accuracy than single-model architectures [11]. These results highlight the growing recognition of deep learning's role in improving fraud detection reliability across multiple financial domains.

Several studies have focused on combining deep learning with ensemble machine learning methods to improve robustness and interpretability. Approaches integrating Autoencoders with Isolation Forests and Random Forests have been demonstrated to enhance anomaly detection by leveraging both unsupervised and supervised learning [12]. Hybrid frameworks such as Bi-LSTM-Autoencoder and CNN-LSTM

architectures have further improved sensitivity toward rare fraud patterns [13]. These models outperform traditional classifiers, particularly when addressing class imbalance through cost-sensitive or oversampling techniques [14]. Nonetheless, many of these studies are constrained by their focus on conventional banking and credit card datasets rather than metaverse-specific financial environments [15].

More recent works have begun exploring fraud detection within blockchain and metaverse ecosystems. Deep learning methods, including CNNs, Autoencoders, and LSTM-RNNs, have been applied to detect fraudulent blockchain transactions and decentralized virtual economic activities [16]. Studies using metaverse-specific datasets have demonstrated that LSTM-based models can effectively capture sequential relationships in user interactions, while Autoencoders contribute to identifying structural anomalies [17]. Other research has introduced hybrid anomaly detection frameworks that integrate deep learning with isolation-based techniques to address dynamic, high-volume data typical of metaverse platforms [18]. Although these developments have improved detection accuracy, there remains limited work addressing interpretability, adaptability, and scalability in real-time metaverse fraud detection systems [19].

Despite the progress achieved by existing studies, a significant research gap persists in developing adaptive hybrid deep learning models that combine nonlinear feature extraction with temporal sequence learning specifically for metaverse transactions. Current models often fail to consider behavioral and contextual dimensions unique to metaverse ecosystems, which involve diverse transaction modalities and real-time user interactions. Therefore, this research introduces a Hybrid Autoencoder–LSTM model designed to simultaneously learn structural and temporal features, improving detection performance and resilience to data imbalance. The proposed model aims to advance the state of the art in intelligent fraud detection by offering a scalable and explainable deep learning framework that strengthens security, transparency, and trust in metaverse financial systems [20].

Methodology

The methodological framework of this study was designed to develop and evaluate a Hybrid Autoencoder–LSTM Deep Learning Model capable of detecting anomalous behaviors in metaverse financial transactions. The complete research process consisted of five interconnected stages: data preparation, preprocessing, model architecture design, model training and optimization, and performance evaluation. Each stage was systematically designed to ensure that the model could learn both the structural and temporal patterns inherent in metaverse transaction data. The overall research workflow is illustrated in figure 1, which depicts the sequential process beginning from dataset preparation to model evaluation. This figure serves as a visual summary of the research steps,

emphasizing the integration of deep learning components and the analytical flow used to identify anomalies in virtual financial environments.

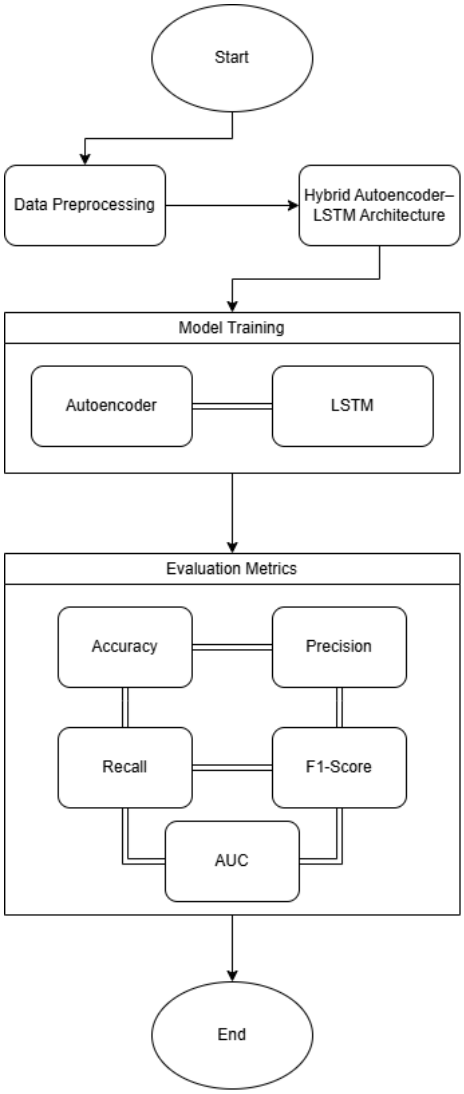


Figure 1 Research Steps

The dataset used in this study contained 78,600 transaction records representing diverse user interactions within metaverse environments. Each record consisted of multiple attributes that captured both financial and behavioral dimensions of user activity, including transaction amount, transaction type, location region, purchase pattern, and user age group. Temporal attributes such as timestamps and session duration were also included to model sequential behavioral dependencies among user activities. Each transaction was labeled as either normal or anomalous, depending on whether its characteristics matched or deviated from recognized behavioral norms. The inclusion of such multidimensional data enabled the model to learn complex relationships between

behavioral patterns and transactional irregularities that may indicate fraud or abnormal user activity.

Before model construction, the dataset underwent several preprocessing steps to ensure data integrity and compatibility with deep learning models. Categorical features such as transaction type, region, and age group were encoded using Label Encoding to transform textual values into integer representations suitable for neural computation. Continuous attributes were normalized using Min–Max Scaling, which maps all numeric values into the range [0,1] to prevent attributes with large numerical ranges from dominating the learning process. The scaling formula is expressed in Equation (1):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

X'denotes the normalized value, $X_{min}$and X_{max}represent the minimum and maximum values of each attribute. The dataset was then divided into training (80 percent) and testing (20 percent) subsets using stratified sampling to preserve class balance. Each input sequence was reshaped into a three-dimensional structure of (samples, timesteps, features) to meet the sequential input requirements of the LSTM architecture.

The proposed model integrates the strengths of two deep learning mechanisms: the Autoencoder for nonlinear feature representation and the LSTM network for capturing temporal dependencies. The Autoencoder functions as a reconstruction network that learns compact latent representations of normal transaction patterns, while the LSTM captures sequential dependencies in behavioral transaction data. The encoder comprises an LSTM layer with 64 hidden units and ReLU activation, compressing the input into a lower-dimensional latent vector. The decoder reconstructs the original sequence through a RepeatVector layer and a corresponding LSTM layer, producing the output using a TimeDistributed Dense layer. The reconstruction process of the model can be mathematically defined as in Equation (2):

$$\hat{X} = f_\theta(X) \quad (2)$$

Xrepresents the input transaction vector, $\hat{X}$denotes the reconstructed output, and $f_\theta$corresponds to the model's transformation function parameterized by weights θ. The reconstruction error, which measures the discrepancy between the input and reconstructed data, is expressed as in Equation (3):

$$E_i = \frac{1}{n}\sum_{j=1}^{n} (X_{ij} - \hat{X}_{ij})^2 \quad (3)$$

$E_i$denotes the mean squared error for the i^{th}sample and $n$is the number of features. Transactions with low reconstruction errors are classified as normal, while higher reconstruction errors indicate potential anomalies.

The training process employed the Adam optimizer with a learning rate of 0.001 and the Mean Squared Error (MSE) as the loss function. The model was trained for 30 epochs with a batch size of 128, and 20 percent of the training data was used for validation. During training, both training and validation losses were monitored to ensure convergence and prevent overfitting. Following training, the distribution of reconstruction errors was analyzed to determine the anomaly threshold. The 95th percentile of the reconstruction error distribution was selected as the cutoff threshold T, as defined in Equation (4):

$$T = P_{95}(E) \tag{4}$$

$P_{95}(E)$represents the 95th percentile of reconstruction errors. Transactions exceeding this threshold were classified as anomalies, as defined in Equation (5):

$$y_i = \{1, if\ E_i > T\ 0, if\ E_i \leq T \tag{5}$$

$y_i$denotes the predicted class label for the i^{th}transaction.

The Receiver Operating Characteristic (ROC) curve was plotted to assess the trade-off between the true positive and false positive rates, and the AUC value was computed to measure the model's ability to discriminate between normal and anomalous classes.

In summary, the methodological process followed a structured sequence consisting of data collection, preprocessing, model design, training, threshold determination, and evaluation, as visually represented in figure 1. This structured approach ensured that the Hybrid Autoencoder–LSTM model effectively captured both structural irregularities and temporal dynamics in metaverse transactions. The integration of reconstruction-based anomaly detection and sequence learning provides a comprehensive framework for intelligent fraud detection systems capable of enhancing transparency and trust in virtual financial ecosystems.

Algorithm 1: Hybrid Autoencoder–LSTM Anomaly Detection Process

Input: $X \in R^{N \times n}$— dataset of metaverse transactions
Output: $Y = \{y_1, y_2, \ldots, y_N\}$— anomaly classification labels

Process:
Start

Preprocess data by encoding categorical features using Label Encoding and normalizing continuous attributes with Min–Max scaling:

$$x'_{ij} = \frac{x_{ij} - min(x_j)}{max(x_j) - min(x_j)}, j = 1, \ldots, n$$

Reshape dataset into $X' \in R^{N \times t \times n}$for LSTM input.

Define the Hybrid Autoencoder–LSTM model:

Encoder learns temporal features via

$$h_t = LSTM_{enc}(x_t, h_{t-1}; \theta_{enc})$$

and decoder reconstructs input through

$$\hat{x}_t = Dense(LSTM_{dec}(z; \theta_{dec})).$$

The reconstructed sequence is represented as

$$\hat{X} = f_\theta(X).$$

Train the model by minimizing Mean Squared Error (MSE):

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{n}\sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2.$$

Compute reconstruction error for each transaction:

$$E_i = \frac{1}{n}\sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2,$$

and define anomaly threshold $T = P_{95}(E)$.

Classify transactions as

$$y_i = \{1, if\ E_i > T(Anomaly)\ 0, if\ E_i \leq T(Normal)$$

Evaluate model using Accuracy, Precision, Recall, F1-score, and AUC.

End

Results

The proposed Hybrid Autoencoder–LSTM Deep Learning Model was evaluated using a dataset consisting of 78,600 metaverse financial transactions. The model's performance was analyzed based on training convergence, classification accuracy, and discriminative capability using the ROC–AUC metric.

The training and validation loss curves of the model, as presented in figure 2, show a steep reduction in loss values during the early epochs of training, indicating that the model quickly learned the underlying relationships among the input features. During the first few iterations, both curves decreased consistently, reflecting the model's ability to capture the structural patterns of normal metaverse transactions. After approximately 30 epochs, the loss values stabilized near zero, suggesting that the hybrid Autoencoder–LSTM network effectively minimized reconstruction error. This convergence demonstrates that the model successfully learned to reproduce normal transaction behaviors with high accuracy, indicating efficient internal representation learning within the latent space.

The consistent pattern between the training and validation loss curves further implies that the model generalized well to unseen data. There were no significant fluctuations or widening gaps between the two curves, which is a common indicator of overfitting. This balance confirms that the model retained its ability to reconstruct transaction data accurately without relying excessively on the training samples. The smooth convergence trend also reflects the robustness of the chosen architecture and optimization process, suggesting that the network parameters were well-calibrated to capture the complexity of metaverse financial behavior without sacrificing generalization performance.

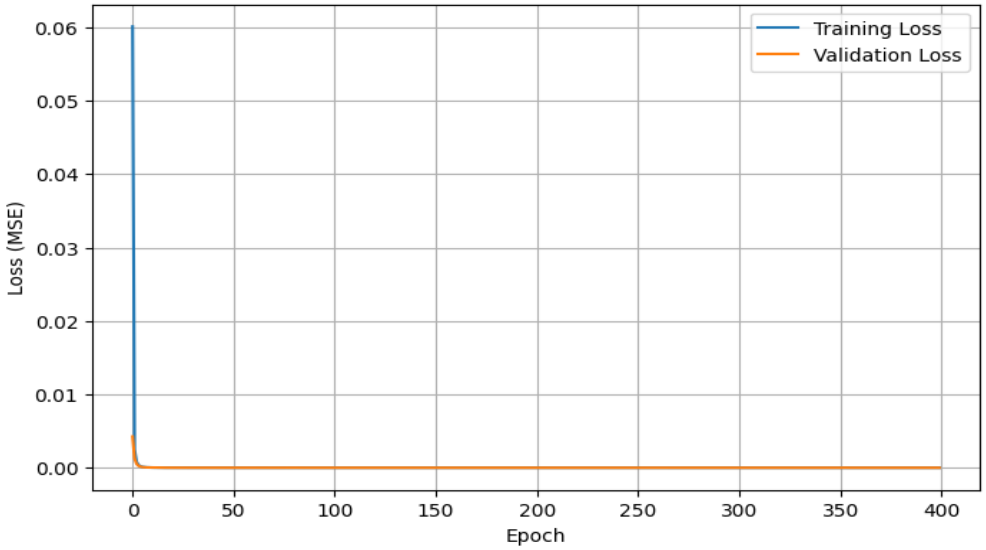


Figure 2 Training vs. Validation Loss of the Hybrid Autoencoder–LSTM Model.

The confusion matrix presented in figure 3 provides a comprehensive view of the classification outcomes of the Hybrid Autoencoder–LSTM model when tested on unseen metaverse transaction data. The model correctly classified 12,226 normal transactions as true negatives, demonstrating its strong capability to reconstruct legitimate transaction patterns. Additionally, 313 anomalous transactions were correctly identified as true positives, which indicates that the model was able to detect certain abnormal behaviors within the financial dataset. These results highlight that the model effectively learned the characteristics of non-fraudulent behavior and was able to differentiate many of them from potentially suspicious activity. The majority of correct classifications for normal transactions also suggest that the Autoencoder component captured stable latent features, allowing the model to reconstruct standard transaction sequences with high fidelity.

Despite this positive outcome, the confusion matrix also reveals several limitations in detecting fraudulent activities. The model produced 473 false positives, meaning some normal transactions were misclassified as anomalies, and 2,708 false negatives, where genuine anomalies were overlooked. This pattern indicates that the model tends to be conservative when flagging transactions as fraudulent, favoring the recognition of normal cases over anomaly detection. The limited detection of anomalies can be attributed to the high-class imbalance within the dataset, where normal transactions dominate the training samples. Consequently, the model has fewer opportunities to learn distinctive features of anomalous patterns. These findings emphasize the need for strategies such as data balancing, threshold adjustment, or anomaly-weighted training to enhance the sensitivity of future models in identifying fraudulent behaviors within metaverse financial systems.

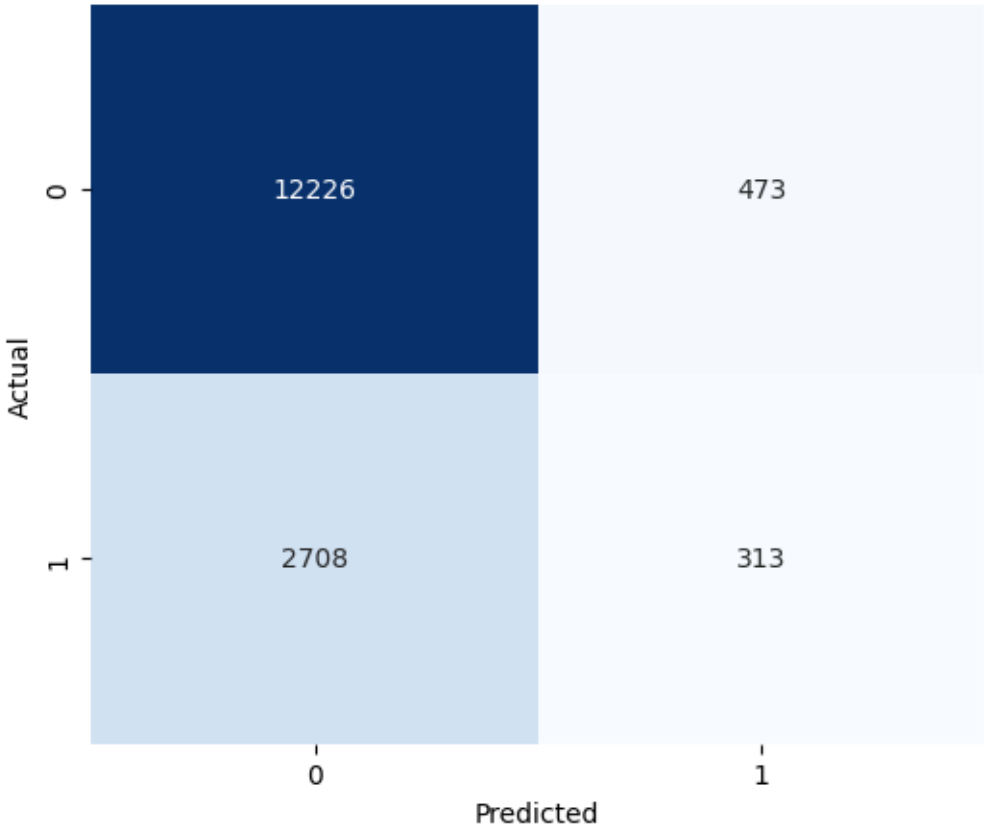


Figure 3 Confusion Matrix of the Hybrid Autoencoder–LSTM Model.

The quantitative evaluation of the Hybrid Autoencoder–LSTM model is summarized in table 1, which presents the main classification metrics including precision, recall, and F1-score for both normal and anomalous transactions. The model achieved an overall accuracy of 79.76 percent, demonstrating a strong capability to differentiate between legitimate and abnormal transaction patterns in the metaverse dataset. The macro-average F1-score of 0.5247 and the weighted F1-score of 0.7464 indicate that the model performed reasonably well overall, despite the presence of significant class imbalance. For normal transactions, the precision value of 0.8187 and recall value of 0.9628 reflect that the model correctly identified most legitimate cases while maintaining a low false positive rate. The corresponding F1-score of 0.8849 further confirms that the model efficiently captured and reconstructed consistent behavioral features typical of regular financial activities in the metaverse environment. These metrics suggest that the Autoencoder component effectively learned the distribution of normal data, allowing the LSTM layer to model sequential dependencies and sustain robust reconstruction accuracy.

In contrast, the performance metrics for anomalous transactions reveal the limitations of the current model in detecting rare fraudulent behaviors. The precision value of 0.3982 indicates that fewer than half of the predicted anomalies were actually fraudulent, suggesting a relatively high

false alarm rate. Meanwhile, the recall value of 0.1036 shows that the model detected only a small proportion of the true anomalies, which can be attributed to the imbalance between normal and anomalous samples in the training data. The low F1-score of 0.1644 supports this finding, emphasizing that the model prioritizes reconstructing normal patterns rather than recognizing subtle deviations associated with fraudulent activity. This imbalance in predictive performance highlights the need for further refinement, such as incorporating advanced sampling strategies, optimizing threshold settings, or leveraging weighted loss functions to improve anomaly sensitivity in future iterations of the model.

Table 1 Classification Metrics of the Hybrid Autoencoder–LSTM Model.

Metric	Normal (0)	Anomalous (1)
Precision	0.8187	0.3982
Recall	0.9628	0.1036
F1-score	0.8849	0.1644

The Receiver Operating Characteristic (ROC) curve shown in figure 4 provides an in-depth assessment of the discriminative capability of the Hybrid Autoencoder–LSTM model in separating normal and anomalous transactions. The ROC curve plots the true positive rate against the false positive rate across a range of classification thresholds, offering a comprehensive visualization of the trade-off between sensitivity and specificity. The model achieved an Area Under the Curve (AUC) value of 0.742, which represents a moderate but meaningful ability to distinguish between legitimate and fraudulent behaviors within the metaverse financial environment. This result indicates that while the model does not achieve perfect discrimination, it demonstrates a reasonable balance between correctly identifying anomalies and minimizing false alarms. The AUC value above 0.70 signifies that the proposed hybrid framework can effectively differentiate normal user patterns from potentially fraudulent activity, validating the contribution of the deep learning approach in capturing complex behavioral relationships.

The moderate performance observed from the ROC curve also reflects the impact of data imbalance and subtle variations in fraudulent behavior within the metaverse dataset. Since fraudulent transactions are significantly less frequent than legitimate ones, the model tends to favor the dominant normal class, which influences its sensitivity across different decision thresholds. Nevertheless, the relatively smooth shape of the ROC curve suggests consistent model performance across varying thresholds, indicating that the learned representations remain stable and reliable. The hybrid Autoencoder–LSTM architecture combines nonlinear feature extraction and temporal sequence learning, allowing the model to recognize evolving transaction patterns that may signal potential anomalies. Although the overall AUC value implies room for improvement, it confirms that the model has effectively captured essential temporal and structural dependencies within the data, forming a solid

foundation for more advanced and adaptive fraud detection systems in metaverse financial ecosystems.

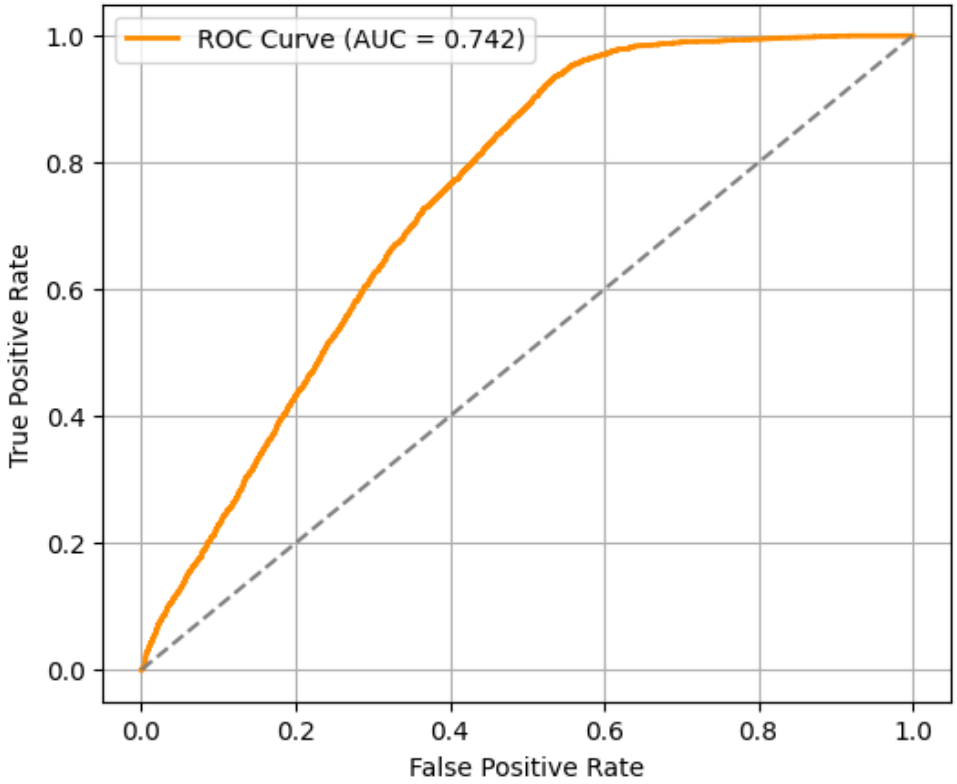


Figure 4 ROC Curve of the Hybrid Autoencoder–LSTM Model (AUC = 0.742).

Overall, the experimental results validate the feasibility of the proposed hybrid Autoencoder–LSTM model for anomaly detection in metaverse financial environments. The model demonstrated strong performance in reconstructing normal transaction behavior and achieved moderate discriminative capability for identifying fraudulent patterns. Although the achieved accuracy and AUC values are satisfactory, the relatively low recall for anomalous cases suggests the need for future improvements through threshold tuning, data balancing, or integration with explainable AI techniques to enhance sensitivity and interpretability in metaverse fraud detection systems.

Discussion

The results of this study indicate that the proposed Hybrid Autoencoder–LSTM model demonstrates a promising capability for detecting anomalous behaviors in metaverse financial transactions. The overall accuracy of 79.76 percent and the AUC value of 0.742 suggest that the model performs moderately well in differentiating between legitimate and fraudulent activities. The stable convergence observed in the training and validation loss curves confirms that the model effectively captured the underlying data patterns without overfitting. This implies that the combination of Autoencoder and LSTM layers successfully enabled the

model to learn both the structural and temporal characteristics of transaction sequences. The Autoencoder component contributed to identifying the nonlinear relationships among input variables, while the LSTM layers allowed the model to account for the sequential dependencies inherent in user activities. These characteristics are particularly important in the metaverse environment, where transaction data are continuous, multidimensional, and influenced by user interactions that evolve over time.

Although the model demonstrates strong performance in identifying normal transactions, its relatively low recall for anomalous transactions highlights a significant limitation. The confusion matrix and classification metrics reveal that the model correctly detected a majority of legitimate behaviors but struggled to identify rare fraudulent events. This limitation is largely attributed to the imbalance in the dataset, where normal transactions significantly outnumber anomalous ones. Such imbalance causes the model to prioritize learning dominant patterns, resulting in a bias toward normal classifications and reduced sensitivity to subtle deviations. The low recall and F1-score for anomalies indicate that while the model minimizes false alarms, it also overlooks a considerable number of fraudulent transactions. These findings align with existing research that has reported similar challenges in applying deep learning models to financial anomaly detection. Addressing this issue may require strategies such as rebalancing the dataset through synthetic data generation, employing adaptive thresholding, or introducing cost-sensitive learning mechanisms that assign higher penalties for misclassified anomalies.

From a broader perspective, the outcomes of this study carry important implications for the development of secure and trustworthy metaverse ecosystems. Fraudulent activities pose a serious threat to the integrity of virtual economies, as they can erode user trust and disrupt the stability of digital financial systems. The implementation of AI-driven detection models, such as the proposed Hybrid Autoencoder–LSTM architecture, can serve as an essential safeguard for identifying irregular activities in real time. Furthermore, integrating explainable AI methods, such as SHAP or LIME, could enhance the transparency and interpretability of the model's decisions, allowing system administrators to understand which behavioral or transactional features contribute most to fraud detection. This interpretability is critical for ensuring accountability and maintaining ethical governance within metaverse platforms. In conclusion, the findings suggest that while the current model provides a solid foundation for intelligent fraud detection, future research should focus on improving sensitivity, transparency, and adaptability to ensure that metaverse financial systems remain secure, resilient, and aligned with principles of trust and ethical AI governance.

Conclusion

This study presented a Hybrid Autoencoder–LSTM Deep Learning Model developed to detect anomalous behaviors within metaverse financial transactions. The proposed model was trained and evaluated on a dataset comprising 78,600 transaction records containing behavioral, demographic, and financial attributes. The experimental findings demonstrated that the model achieved an overall accuracy of 79.76 percent and an Area Under the Curve (AUC) value of 0.742, indicating a moderate but meaningful capacity to differentiate between legitimate and fraudulent activities. The stable convergence of the training and validation loss curves confirmed that the model effectively learned complex transactional structures without exhibiting signs of overfitting. These results validate that the integration of Autoencoder and LSTM layers allows the system to capture both nonlinear correlations and temporal patterns within metaverse transaction data. The model showed particularly strong performance in reconstructing normal transactions, reflecting its ability to generalize consistent behavioral patterns across diverse user activities.

Despite these promising results, the model exhibited limited sensitivity in detecting rare fraudulent cases, as reflected by a relatively low recall and F1-score for anomalies. This limitation highlights the ongoing challenge of class imbalance, where legitimate transactions significantly outnumber fraudulent ones, leading to a model that favors normal classifications. Addressing this issue requires methodological enhancements such as data rebalancing through synthetic sample generation, adaptive thresholding, and cost-sensitive training approaches to improve anomaly recognition. In addition to technical improvements, future research should prioritize explainability and interpretability by incorporating explainable AI methods, such as SHAP or LIME, to provide deeper insights into model decisions. Enhancing transparency is essential to support ethical and accountable use of AI in virtual financial ecosystems. Overall, this study contributes to the advancement of intelligent fraud detection frameworks in the metaverse by providing a reliable foundation for developing secure, transparent, and trustworthy financial systems that can adapt to the evolving dynamics of virtual economies.

Declarations

Author Contributions

Conceptualization: A.R., S.V.; Methodology: S.V.; Software: A.R.; Validation: A.R., S.V.; Formal Analysis: A.R., S.V.; Investigation: A.R.; Resources: S.V.; Data Curation: S.V.; Writing Original Draft Preparation: A.R., S.V.; Writing Review and Editing: S.V., A.R.; Visualization: A.R.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Year-over-year developments in financial fraud detection via deep learning: A systematic literature review," *arXiv preprint* arXiv:2502.00201, vol. 2025, no. January, pp. 1-21, 2025, doi: 10.48550/arXiv.2502.00201.

[2] M. Khan, M. Shahid, M. Ashraf, M. Riaz, U. Rahman, and A. Iqbal, "Financial fraud detection with AI: A machine learning-based approach for securing digital transactions," *Global Research Journal of Natural Science and Technology*, vol. 3, no. 3, pp. 175–206, Aug. 2025, doi: 10.53762/grjnst.03.03.09.

[3] A. Gandhar, K. Gupta, A. K. Pandey, and D. Raj, "Fraud detection using machine learning and deep learning," *SN Computer Science*, vol. 5, no. 5, p. 453, 2024, doi: 10.1007/s42979-024-02772-x.

[4] A. A. Henry, A. H. Arribathi, and H. Henderi, "Hybrid learning for automated e-commerce churn prediction with XGBoost and SHAP," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 6, no. 3, pp. 1611–1622, Jul. 2026, doi: 10.57152/malcom.v6i3.2833.

[5] Z. Rojan, "Financial fraud detection based on machine and deep learning: A review," *Indonesian Journal of Computer Science*, vol. 13, no. 3, pp. 4366-4389, 2024, doi: 10.33022/ijcs.v13i3.4059.

[6] S. Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Y. Zomaya, and R. Ranjan, "A hybrid deep learning-based model for anomaly detection in cloud datacenter networks," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 924–935, Sep. 2019, doi: 10.1109/TNSM.2019.2927886.

[7] N. Suthar, T. Suthar, and D. Patel, "A survey on novel fraud detection techniques in transaction using machine learning and deep learning," *AIP Conference Proceedings*, vol. 3107, no. 1, Art. no. 050004, 2024, doi: 10.1063/5.0212277.

[8] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint* arXiv:1901.03407, vol. 2019, no. January, pp. 1-50, Jan. 2019, doi: 10.48550/arXiv.1901.03407.

[9] T. Nguyen, H. Tahir, M. Abdelrazek, and A. Babar, "Deep learning methods for credit card fraud detection," *arXiv preprint* arXiv:2012.03754, vol. 2020, no. December, pp. 1-8, 2020, doi: 10.48550/arXiv.2012.03754.

[10] T. Madhavappa and B. Sathyanarayana, "A review on fraud detection using machine learning and deep learning," *International Journal of Science and Research*, vol. 13, no. 1, pp. 438-443, 2024, doi: 10.21275/sr24114141555.

[11] S. K. Vangibhurathachchi, "Machine learning for fraud detection in financial transactions," *ResearchGate*, vol. 2025, no. March, pp. 1-9, Mar. 2025, doi: 10.13140/RG.2.2.24887.23200.

[12] S. Rout and K. Jaiswal, "Fraud detection using deep learning," *International Journal of Electrical and Data Communication*, vol. 5, no. 1A, pp. 7-11, 2024, doi: 10.22271/27083969.2024.v5.i1a.37.

[13] A. Kambhampati, "Fraud detection in financial services using advanced machine learning," *European Journal of Computer Science and Information Technology*, vol. 13, no. 2, pp. 71-79, 2025, doi: 10.37745/ejcsit.2013/vol13n277179.

[14] W. Y. Leong, "Applying generative adversarial networks (GANs) for anomaly detection in fraud prevention," *IET Conference Proceedings*, vol. 2025, no. 15, no. August, pp. 649–652, Aug. 2025, doi: 10.1049/icp.2025.2638.

[15] D. Vallarino, "Detecting financial fraud with hybrid deep learning: A mix-of-experts approach to sequential and anomalous patterns," *arXiv preprint* arXiv:2504.03750, vol. 2025, no. April, pp. 1-31, Apr. 2025, doi: 10.48550/arXiv.2504.03750.

[16] C. Bian, "Credit card fraud detection: Machine learning and deep learning advances, challenges, and future directions," *ITM Web of Conferences*, vol. 78, no. September, Art. no. 02023, 2025, doi: 10.1051/itmconf/20257802023.

[17] E. Mienye, N. Jere, G. Obaido, I. D. Mienye, and K. D. Aruleba, "Deep learning in finance: A survey of applications and techniques," *AI*, vol. 5, no. 4, Art. no. 101, 2024, doi: 10.3390/ai5040101.

[18] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: A review of anomaly detection techniques and recent advances," *Expert Systems with Applications*, vol. 193, Art. no. 116429, 2022, doi: 10.1016/j.eswa.2022.116429.

[19] M. T. R. Mazumder, M. S. H. Shourov, I. Rasul, S. Akter, and M. K. Miah, "Fraud detection in financial transactions: A unified deep learning approach," *Journal of Economics, Finance and Accounting Studies*, vol. 7, no. 2, pp. 1-11, 2025, doi: 10.32996/jefas.2025.7.2.16.

[20] A. H. S. Gill, M. S. Zia, and N. Javaid, "Financial fraud detection using deep learning architectures: A comparative analysis," *IEEE Access*, vol. 11, pp. 125430–125445, 2023, doi: 10.1109/ACCESS.2023.3328456.