JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Hersyaputra, M. S., Jaya, M. T. T., & Anggraini, R. N. E. (2026). A hybrid machine learning model for streaming frequency-based anomaly detection in banking transactions. *Journal of Information and Communication Technology, 25*(1), 64-78. https://doi.org/10.32890/jict2026.25.1.4

A Hybrid Machine Learning Model for Streaming Frequency-Based Anomaly Detection in Banking Transactions

[1]Mohamad Syazimmi Hersyaputra, [2]Muhammad Triyanda Taruna Jaya & [3]Ratih Nur Esti Anggraini
[1,2&3]Department of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia

[1]6025241017@student.its.ac.id
[2]6025241014@student.its.ac.id
*[3]ratih@its.ac.id
**Corresponding author*

Received: 1/7/2025 Revised: 1/10/2025 Accepted: 25/10/2025 Published: 31/1/2026

ABSTRACT

The increasing sophistication of fraudulent activities in digital financial systems necessitates real-time anomaly detection models that can adapt to evolving transactional behaviours. While prior research has explored machine learning approaches for fraud detection, most rely on static datasets and overlook temporal dependencies inherent in financial transaction streams. This study addresses this critical gap by proposing a hybrid anomaly detection model that integrates dynamic streaming-frequency analysis via a 7-day sliding window, an unsupervised Long Short-Term Memory (LSTM) -autoencoder for anomaly scoring, and a supervised Artificial Neural Network (ANN) for classification. The sliding window mechanism enables the model to capture short-term temporal fluctuations and behavioural patterns, aligning with the streaming nature of financial data. The LSTM-autoencoder is trained exclusively on normal transaction sequences to learn temporal dependencies and compute reconstruction errors, which serve as deep anomaly features. These features are then fed into the ANN to classify transactions as normal or anomalous. Experimental results on the IBM Anti-Money Laundering (AML) dataset demonstrate the effectiveness of the proposed framework, achieving a classification accuracy of 99.92%, precision of 94.12%, recall of 88.43%, and F1 score of 91.19%. This layered architecture not only enhances early detection of anomalous behaviour but also provides a scalable, adaptive solution for real-time fraud detection in streaming financial environments.

Keywords: Artificial neural network, financial anomaly detection, LSTM-autoencoder, sliding window, streaming frequency.

INTRODUCTION

Anomaly detection has become a significant challenge, drawing increasing attention across various research domains and real-world applications (Anggraini et al., 2023). This interest is further amplified by advancements in Information Technology (IT), which have transformed modern industrial society, particularly through the rise of internet-based financial technologies (Koo et al., 2024). With the rapid development of the digital era, safeguarding the security and integrity of financial transactions has become a critical priority. Although the shift to digital platforms provides unmatched ease and accessibility for users, it simultaneously introduces new vulnerabilities, most notably, the risk of banking transaction fraud (Cherif et al., 2023). These cyber threats, especially in the context of online banking, have far-reaching consequences (Awosika et al., 2024). They not only undermine user confidence and disrupt personal financial stability but also threaten the credibility of financial institutions and pose substantial risks to the broader national economy (Hilal et al., 2021).

Traditional methods of detecting suspicious transactions, typically reliant on rule-based systems and manual reviews, face significant challenges, including inefficiency, high false-positive rates (Irshad et al., 2024), and limited adaptability to emerging fraud patterns (Ketenci et al., 2021). Traditional anomaly detection methods also may fall short in capturing dynamic transactional patterns in a streaming data environment (Ketenci et al., 2021). Manual reviews for detecting fraudulent transactions are often inefficient and resource-intensive, prompting a growing reliance on machine learning techniques that have increasingly supplanted traditional methods (Sehrawat & Singh, 2023). To overcome these limitations, advanced detection techniques leveraging machine learning and data mining have emerged, capable of analysing large transaction datasets to automatically identify unusual patterns and anomalies (Kute et al., 2021).

Current research on detecting suspicious financial transactions increasingly leverages machine learning algorithms due to their ability to learn complex patterns and adapt to evolving fraud behaviours (Ghimire, 2025). Among the commonly used models, Random Forest stands out for its robustness and ability to handle imbalanced datasets through ensemble learning, making it effective in distinguishing rare fraudulent activities from legitimate ones (Paldino et al., 2024). Support Vector Machines (SVMs) are valued for their high accuracy in binary classification problems and their effectiveness in detecting fraud with clear margins between classes (Dastidar, 2024). Artificial Neural Networks (ANN), particularly deep learning variants, can model intricate non-linear relationships within large transaction datasets, enabling them to uncover subtle fraud patterns (Almazroi & Ayub, 2023).

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, are particularly powerful in capturing temporal dependencies within transaction sequences, allowing them to detect fraudulent behaviour that unfolds over time or through sequential patterns (Ileberi & Sun, 2024). XGBoost has gained popularity for its scalability, fast execution, and superior performance in structured data scenarios, often outperforming traditional methods in fraud prediction tasks (Ileberi & Sun, 2024). Additionally, autoencoders, an unsupervised deep learning technique, are increasingly used for anomaly detection by learning to reconstruct normal behaviour and flagging deviations as potential fraud (Alarfaj & Shahzadi, 2025). Each of these models offers unique advantages, contributing to a more comprehensive and adaptive fraud-detection framework in financial systems.

Recent studies have further highlighted the importance of combining deep learning with contextual risk-based assessment, as demonstrated by Koo et al. (2024), who integrated autoencoder models with behavioural risk features to strengthen anti-money laundering operations, showing that unsupervised

models can capture hidden patterns beyond the reach of traditional approaches. Similarly, Iqbal and Amin (2024) emphasised the value of LSTM-autoencoders and dimensionality reduction for effectively modelling normal transactional behaviour, while also reviewing hybrid and ensemble methods combining convolutional neural network (CNN), transformer, and generative adversarial network (GAN) architectures, underscoring the growing trend toward architectures that balance accuracy, scalability, and robustness.

Despite the growing adoption of machine learning for detecting suspicious financial transactions, a critical gap remains in both research and practical implementation, particularly in integrating dynamic streaming-frequency analysis with advanced machine learning models. Most existing approaches depend heavily on static datasets or batch-based processing, which are limited in their ability to reflect real-time transactional behaviour and adapt to evolving fraud tactics. In contrast, sliding window techniques offer a dynamic approach by segmenting time series data into overlapping intervals and shifting these windows forward incrementally (Zhang et al., 2020). It enables continuous, context-aware learning from the most recent activity, preserving the temporal structure of the data.

By capturing short-term trends and shifts in behaviour, sliding windows are well-suited for uncovering subtle or gradual anomalies that may not be evident in isolated transactions. When integrated with deep learning models, the system can detect deviations from learned normal patterns in near real time (Farheen & Kumar, 2025). This combination enhances the system's adaptability, enables early anomaly detection, and improves responsiveness to emerging fraud strategies. As a result, the fusion of sliding-window-based temporal modelling with advanced machine learning holds great promise for developing scalable, responsive, and real-time fraud detection systems that align with the dynamic nature of modern financial ecosystems.

To address existing gaps and challenges in financial suspicious transaction detection, this research proposes a hybrid anomaly detection model that integrates both unsupervised and supervised learning components. The approach combines 7-day sliding-window-based streaming frequency analysis for temporal segmentation with an LSTM-autoencoder to model and reconstruct normal transaction behaviour. The sliding window technique allows the model to continuously monitor and capture dynamic behavioural patterns over time, aligning well with the real-time nature of financial transaction streams. The LSTM-autoencoder, trained exclusively on normal transaction sequences, learns temporal dependencies and identifies deviations via reconstruction error, effectively capturing subtle, sequential anomalies.

To enhance classification, the LSTM-autoencoder's reconstruction error is combined with deep temporal features and fed into an ANN, which serves as the supervised layer that classifies transactions as normal or suspicious. This hybrid architecture effectively integrates the strengths of both unsupervised and supervised learning paradigms. Unsupervised learning enables the system to analyse large-scale, unlabeled transaction data and uncover hidden patterns or anomalies that may signify fraudulent behaviour (Yu et al., 2024). This layered architecture represents a significant advancement in fraud detection by leveraging the pattern discovery capabilities of unsupervised learning and the decision-making precision of supervised classification (Faisal et al., 2024). As a result, the model is not only highly adaptive to emerging fraud strategies but also robust in managing imbalanced and heterogeneous financial datasets. Overall, this approach offers a scalable, intelligent, and real-time solution for detecting anomalies in dynamic financial environments.

RELATED WORKS

Suspicious transaction detection in banking has evolved with the adoption of machine learning models that address the limitations of traditional rule-based systems. Two significant directions in the literature are the use of an autoencoder for anomaly detection in financial transactions and the application of time-frequency analysis for pattern recognition in transaction behaviours. Koo et al. (2024) proposed a suspicious financial transaction detection model that integrates an autoencoder with a risk-based approach (RBA) to enhance internal control in Anti-Money Laundering (AML) systems. Their model leverages unsupervised deep neural networks, particularly an autoencoder, to reconstruct transaction data and identify anomalies based on reconstruction errors. This approach eliminates the need for extensive labelled data and is particularly effective at detecting complex, hidden patterns in large, imbalanced datasets. The study emphasises the significance of combining behavioural risk features, such as transaction time, device, and user location, with Autoencoder outputs to quantify the risk of each transaction. After fine-tuning the autoencoder with the RBA framework, the proposed method achieved an accuracy of 99.71%, demonstrating substantially higher performance than traditional AML methods, such as Random Forest, which achieved only 95.13%–95.93%.

Ketenci et al. (2021), on the other hand, introduced a novel AML model utilising time-frequency analysis with a sliding window mechanism to convert financial transactions into 2D spectrogram representations. They applied a Short-Time Fourier Transform (STFT) with a quarterly sliding time window (moved forward daily) to extract behavioural features over both time and frequency domains. This approach captured subtle irregularities in user transaction patterns that are often missed in raw time series data. Using Random Forest and simulated annealing for hyperparameter tuning, the model achieved a notable F1 score and reduced the false-positive rate when time-frequency features were combined with customer relationship management (CRM) data. Specifically, time-frequency features alone yielded a false positive rate of 14.9% and an F1 score of 59.05%, whereas combining them with transaction and CRM features reduced the false positive rate to 11.85% and improved the F1 score to 74.06%. These enhancements substantially improved the area under the curve results by more than 1% compared to existing data science-based transaction monitoring systems.

Recent studies have highlighted the growing importance of deep learning approaches in time series forecasting and anomaly detection. Traditional statistical techniques often fall short in capturing the complexity of modern, high-dimensional, and non-linear financial data. In contrast, deep learning models, particularly LSTMs, autoencoders, GANs, and transformers, have demonstrated superior performance by learning temporal dependencies and intricate patterns in sequential data. Iqbal and Amin (2024) emphasised the effectiveness of LSTM-autoencoders in modelling normal transaction behaviour and identifying anomalies via reconstruction error, and noted improved results when combining LSTM with dimensionality reduction techniques. In their experiments, the LSTM-autoencoder achieved 99.95% accuracy and an F1 score of 98.61% on the Credit Card Fraud Detection dataset, highlighting its superior ability to handle highly imbalanced data. Additionally, the study reviews a variety of hybrid and ensemble methods, including CNN-LSTM, transformer-based architectures, and GANs, which enhance detection accuracy and robustness across different datasets.

To further enhance temporal modelling, Ileberi and Sun (2024) proposed a hybrid deep learning ensemble model for fraud detection that combines CNN, LSTM, and transformer architectures within a stacking framework. The LSTM component effectively captures sequential transaction behaviour and long-term dependencies, critical for identifying delayed or staged fraudulent activities. The model uses XGBoost as a meta-learner to fuse insights from spatial (i.e., CNN), temporal (i.e., LSTM), and

attention-based (i.e., transformer) submodels. The experimental results demonstrate that the hybrid ensemble significantly outperforms individual base learners and traditional methods, achieving a sensitivity of 0.961, specificity of 0.999, and an area under the receiver operating characteristic curve (AUC-ROC) of 0.972 on the European Credit Card Dataset. These findings validate the model's robustness and adaptability.

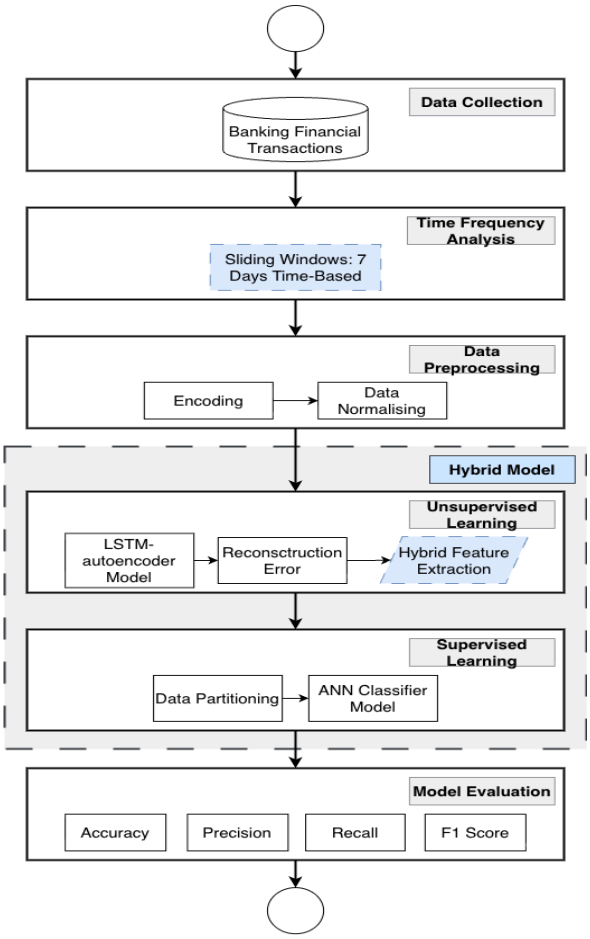
These studies underscore the effectiveness of unsupervised learning (i.e., autoencoders), advanced feature extraction (i.e., time-frequency analysis), and hybrid deep learning architectures (e.g., LSTM-based ensembles) for identifying anomalous transactions in massive financial datasets. However, these approaches have often been applied independently, with autoencoders focusing on latent pattern reconstruction, time-frequency methods capturing dynamic behavioural shifts, and hybrid deep models emphasising sequential learning. This research aims to bridge these dimensions by integrating streaming frequency-based features into a deep sequential framework, such as LSTM or LSTM-autoencoder hybrids, to enable more robust and real-time anomaly detection in banking transactions.

METHODOLOGY

The proposed contributions are explicitly highlighted in Figure 1 with dashed boxes: the use of a sliding-window mechanism for dynamic temporal segmentation and the design of a hybrid model that integrates unsupervised (i.e., LSTM-autoencoder) and supervised (i.e., ANN) learning to enhance anomaly detection in banking transactions.

Figure 1

Research Methodology

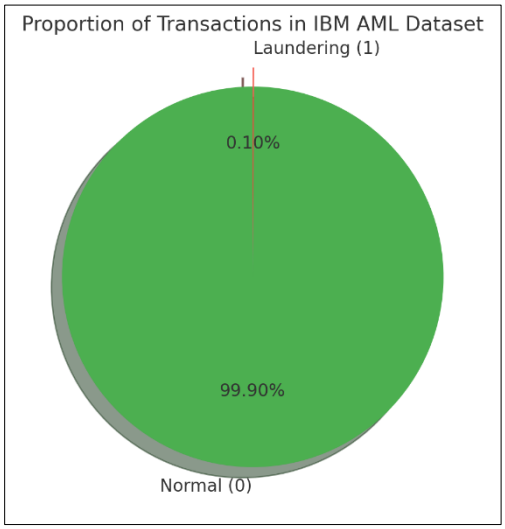


Data Collection

The first phase involves data collection, during which transactional data is sourced from the publicly available IBM AML Transaction Dataset, available on Kaggle. This study explicitly utilises the HI-Small_Trans.csv file, which contains transaction records from Group HI, a customer segment with a relatively higher illicit ratio, indicating a greater prevalence of laundering activity. The dataset contains a total of 11 attributes, which include categorical features stored as objects (i.e., account identifiers, currencies, and payment formats), numerical features represented as integers and floats (i.e., bank codes and transaction amounts), and a binary target variable (i.e., Is Laundering) used to label whether a transaction is suspicious. This file lists approximately 5 million transactions, consisting of 5,073,168 normal records and 5,177 laundering records, as illustrated in Figure 2. The chart highlights the highly imbalanced nature of the target variable, with laundering cases accounting for only a small fraction of total transactions. This labelled dataset provides a suitable basis for training and evaluating both unsupervised and supervised learning models for anomaly detection in financial transactions.

Figure 2

Distribution of Target Variable

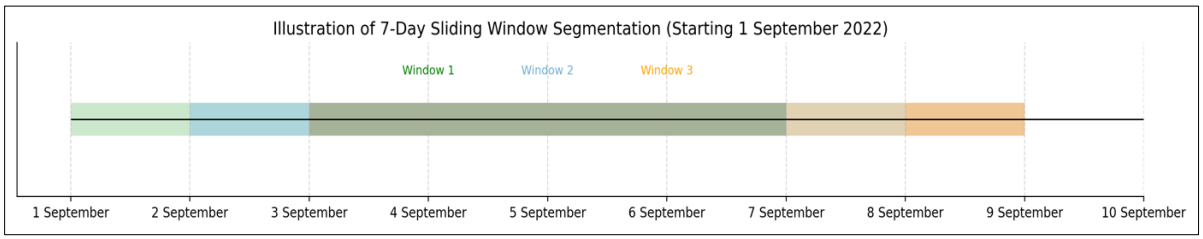


Time Frequency Analysis

The second phase is time-frequency analysis, in which transaction sequences are segmented using a sliding window. In this study, a fixed 7-day window is applied to form temporal segments of user or account activities, meaning that each window groups all transactions occurring within a consecutive 7-day period. The dataset begins on 1 September 2022, so the first window covers transactions from 1 to 7 September 2022; the second window shifts by one day to include transactions from 2 September 2022 to 8 September 2022, and the process continues until the end of the dataset. This overlapping segmentation allows the model to capture short-term behavioural trends and detect anomalies that emerge within specific time intervals, while also maintaining the temporal continuity of transaction flows. Figure 3 illustrates the sliding window mechanism, where each coloured block represents a 7-day segment, and consecutive windows overlap by 6 days to ensure continuous temporal coverage. By applying this procedure, temporal dependencies and frequency-based behaviours are preserved and effectively prepared for sequence modelling in subsequent stages.

Figure 3

Sliding Window Illustration



Data Preprocessing

Once the data is segmented, the next step is data preprocessing, which consists of two key transformations: encoding and normalisation. Categorical variables (i.e., Receiving Currency, Payment Currency, and Payment Format) are encoded using label encoding to convert them to numerical values. Once the data is segmented, the next step is data preprocessing, which consists of two key transformations: encoding and normalisation. Categorical variables are encoded using label encoding techniques to convert them into a numerical format. Following this, data normalisation is performed using a feature-wise standardisation technique to ensure that features with larger numerical ranges do not dominate the training process. Specifically, for each feature x across all time steps in the 3D sliding window sequences X_{seq}, normalisation is defined in Equation 1:

$$x_{scaled} = \frac{x-\mu}{\sigma} \quad (1)$$

where μ is the mean and σ is the standard deviation of the feature across the sequence dimension. The result is a normalised 3D array x_{scaled}, where each individual feature has zero mean and unit variance. This normalisation is crucial for stabilising and accelerating the training of the unsupervised LSTM-autoencoder model. The result is a normalised 3D array X_scaled, where each individual feature has zero mean and unit variance, ensuring that features with larger numerical ranges do not dominate the training process. This normalisation is crucial for stabilising and accelerating the training of the unsupervised LSTM-autoencoder model.

Unsupervised Learning

Following preprocessing, the core of the methodology lies in constructing a hybrid model, initiated by an unsupervised LSTM autoencoder. The LSTM-autoencoder is trained exclusively on normal transaction sequences to reconstruct the input time series. Since the model is optimised to capture only typical patterns, it produces high reconstruction errors when presented with abnormal or suspicious sequences. These reconstruction errors are then interpreted as anomaly scores that quantify the degree of abnormality within a given transaction window. This anomaly score is treated as a hybrid feature, extracted for each transaction segment.

In this study, an LSTM-autoencoder is employed to learn typical patterns of normal financial transactions, enabling the detection of anomalies based on reconstruction errors. One strength of LSTM lies in its memory block, which enables the model to retain and prioritise the most important information (Wardhani et al., 2025). This capability makes LSTM highly effective for modelling complex data that

exhibits long-term dependencies. The model is trained exclusively on sequences labelled as normal to ensure it generalises only legitimate behaviour. The architecture begins with an LSTM encoder layer (LSTM(64, activation = ‘relu’)) that captures temporal dependencies and compresses the input sequence into a latent representation. It is followed by a RepeatVector layer that expands the encoded representation back across the original sequence length, allowing the decoder, another LSTM layer (LSTM(64, activation = ‘relu', return_sequences = True), to reconstruct the input. The output is generated through a TimeDistributed(Dense(n_features)) layer, which produces a feature-wise output at each time step.

The model is compiled using the Adam optimiser for efficient gradient-based optimisation and uses Mean Squared Error (MSE) as the loss function to quantify the difference between original and reconstructed sequences. The training configuration is summarised in Table 1. This configuration ensures stable convergence during training, helps monitor generalisation performance, and prevents overfitting while saving computational resources.

Table 1

Training Configuration for the LSTM-autoencoder Model

Parameter	Value/Setting
Optimiser	Adam
Loss Function	MSE
Epochs	15
Batch Size	64 sequences
Validation Split	0.2 (20% of training data for validation)
Early Stopping	Monitor = 'loss', Patience = 3

The LSTM-autoencoder's output, specifically the reconstruction error, serves as a crucial component of hybrid deep feature extraction in this study. After training the LSTM-autoencoder solely on normal transaction sequences, the model is applied to both normal and anomalous data to compute reconstruction errors for each sequence. These errors, which represent the deviation between the original and reconstructed inputs, act as high-level anomaly scores that capture temporal irregularities in transaction behaviour. Rather than relying solely on raw transactional features, these reconstruction errors are used as enriched features that encapsulate underlying sequential patterns learned by the LSTM architecture. This hybrid representation is then used as the input for a supervised ANN classifier.

Supervised Learning

The hybrid features, particularly the reconstruction errors, are then passed into a supervised learning model, specifically an ANN. Before training the ANN, the dataset is split into training and testing sets of 20% using standard data partitioning techniques. The ANN is trained to classify transactions as either suspicious or normal based on the anomaly score and any additional contextual features. The supervised model complements the unsupervised learning by refining the decision boundary using labelled data, thereby reducing false positives and enhancing overall classification accuracy. The ANN is implemented using a feedforward architecture with multiple hidden layers. The input layer receives the final feature set, which includes reconstruction-based anomaly scores and potentially other contextual features. The model architecture consists of three dense layers with 256, 128, and 64 neurons,

respectively, each followed by a Rectified Linear Unit (ReLU) activation function and a dropout layer (with dropout rates of 0.3, 0.2, and 0.1) to mitigate overfitting. The output layer contains a single neuron with a sigmoid activation function, appropriate for binary classification (i.e., normal vs. anomaly).

The model is compiled with the Adam optimiser (learning rate = 0.001) and binary cross-entropy as the loss function, given the binary nature of the target variable. Accuracy is tracked as the primary evaluation metric during training. To improve generalisation and prevent overfitting, EarlyStopping is employed with a patience value of 5, monitoring the validation loss. The model is trained for up to 30 epochs with a batch size of 64, and 20% of the training data is reserved for validation (validation_split = 0.2). Upon detecting no improvement in validation loss over consecutive epochs, the training halts early, and the best model weights are restored.

Model Evaluation

Finally, the model is subjected to a comprehensive evaluation phase. To assess the effectiveness of the hybrid model approach, the dataset is split into training and test sets, with the test set containing unseen data not used during model development. The model's predictions on this unseen data are then compared against the ground truth labels to evaluate its generalisation capability. Performance metrics such as accuracy, precision, recall, and F1 score are computed. Accuracy provides an overall measure of correctness, while precision and recall offer deeper insights into fraud detection capability, which is particularly important in imbalanced datasets. The F1 score balances precision and recall, serving as a reliable performance indicator in real-world financial scenarios where the costs of false positives and false negatives are asymmetric. This evaluation ensures that the proposed model is not only optimised for training data but also demonstrates robustness and reliability when applied to new, unseen transactions.

RESULTS AND DISCUSSIONS

The proposed method is a hybrid anomaly detection model that integrates a time-frequency-based sliding-window approach, an LSTM-autoencoder for unsupervised anomaly scoring, and an ANN classifier for supervised classification. At this stage, the model's effectiveness in detecting suspicious transactions is evaluated using the IBM AML dataset. The results are analysed in terms of classification performance metrics, confusion matrix interpretation, and the impact of key methodological components such as the sliding window mechanism and the LSTM-autoencoder architecture. This discussion highlights the model's strengths, its ability to handle imbalanced data, and its practical relevance for real-world financial anomaly detection systems.

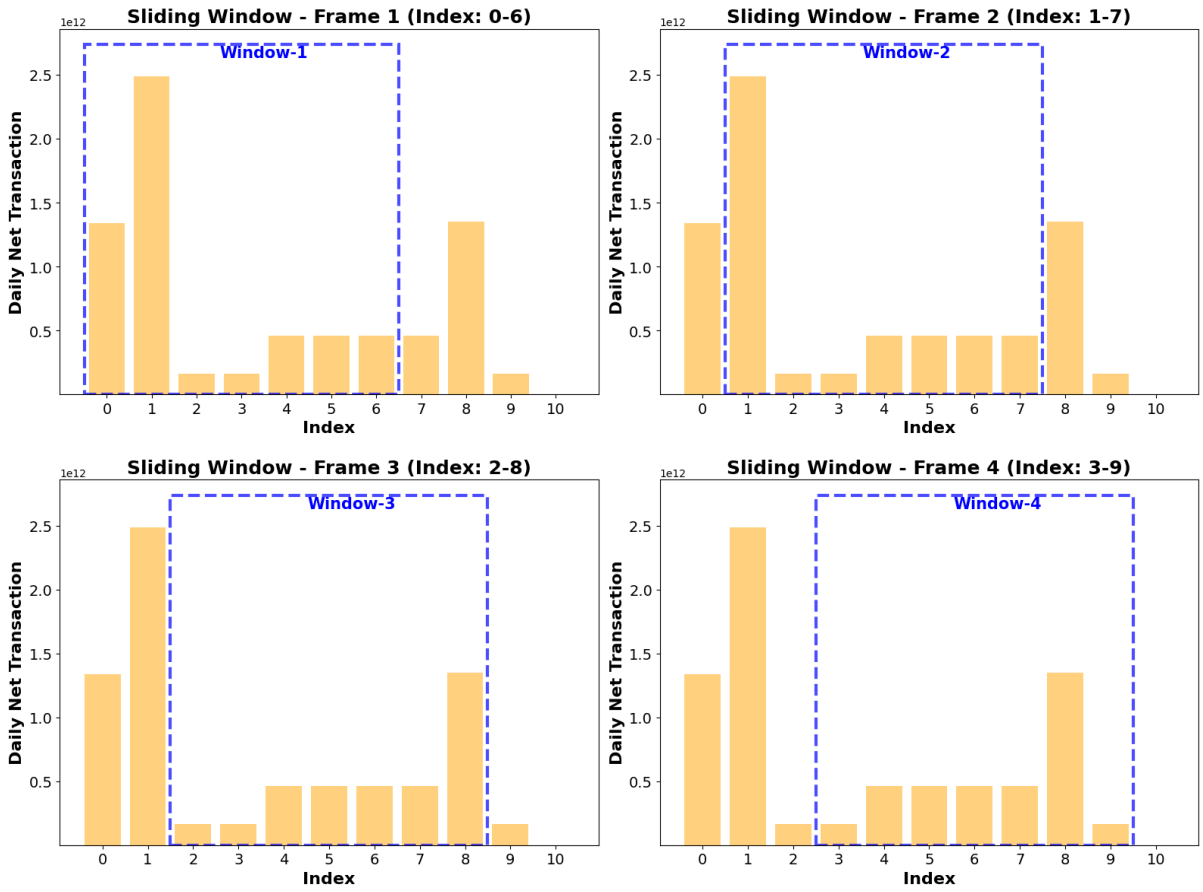
Sliding Window Technique

Applying a 7-day sliding window to the transaction dataset yielded 5,078,338 sequential samples, each represented as a 3D array of shape (7, 10), corresponding to 7 time steps and 10 features per transaction. Each window was assigned a label based on the transaction class (i.e., Is Laundering) at the final time step, forming a label vector of shape (5,078,338). This transformation enabled the dataset to be structured for temporal modelling using sequence-based neural networks. Visual inspection of the first few windows confirmed that the sequences were correctly formed, with each window capturing meaningful transaction behaviour over time. This structured representation enabled the subsequent LSTM autoencoder to effectively learn temporal patterns leading up to either normal or suspicious activity.

The image in Figure 4 illustrates the sliding window technique for temporal segmentation of financial transaction data, specifically focusing on daily net transaction values. In this method, a fixed-size window spanning 7 consecutive indices is moved incrementally across the time series to produce overlapping sequences for model input. Each subplot in the figure represents a different frame (Window-1 to Window-4), with annotations indicating the specific range of indices included in each window. For example, Frame 1 covers indices 0 to 6, which correspond to the first 7-day sliding window of transactions, while Frame 2 spans indices 1 to 7, representing the second 7-day segment that overlaps with the previous window, and so forth. This approach enables the model to observe and learn transaction behaviour patterns over a continuous 7-day period, helping detect trends, sudden changes, and unusual spikes in net transaction volumes. By converting raw transactional data into temporally aware sequences, the sliding window technique allows sequential models to identify better anomalies that may emerge over time, rather than within single-point transactions.

Figure 4

Sliding Window 7 Days Illustration



Effectiveness of the LSTM-autoencoder

To detect anomalies in transaction sequences, an LSTM-autoencoder was trained using windows of normal transactions as a baseline. Each sliding window is assigned a binary label: 0 for normal behaviour and 1 for laundering activity, based on the dataset's ground truth. Stable daily net transaction values characterise normal windows with no unusual spikes or abrupt changes, while anomalous

windows are marked by irregular patterns such as sudden increases in transaction amounts, sharp declines, or inconsistent fluctuations across the 7-day segments. These distinctions are illustrated in Figure 3, where sequences labelled 0 show consistent, stable patterns, whereas sequences labelled 1 highlight abnormal behaviours suggestive of money laundering. The model architecture consists of an input layer that accepts 7-time steps with 10 features each, followed by a 64-unit LSTM-encoder, a RepeatVector layer, a 64-unit LSTM decoder, and a TimeDistributed dense layer to reconstruct the original input. The total number of trainable parameters is 52,874. Training was conducted for up to 15 epochs with early stopping (patience = 3), a batch size of 64, and a 20% validation split. The model converged in under 10 epochs, achieving a final validation loss of 0.1420, showing that it could accurately learn and reconstruct normal transaction patterns with minimal overfitting.

After training, the encoder was used to extract latent representations (i.e., X_encoded) from all sequences, producing vectors of shape (5078338, 64). In parallel, the reconstruction error, calculated as the mean-squared difference between the original and reconstructed sequences, was computed for each window, yielding one anomaly score per sequence. These scores formed a vector of shape (5078338). The encoded representations and reconstruction errors were then combined into a single hybrid feature matrix (X_features) with a shape of (5078338, 65). The LSTM autoencoder proved effective at capturing temporal behaviour and identifying deviations. High reconstruction errors indicated unusual patterns and were used as key features in the supervised classification phase. This unsupervised feature extraction stage provided a strong foundation for the downstream ANN classifier, improving its ability to detect anomalies even without relying heavily on labelled data.

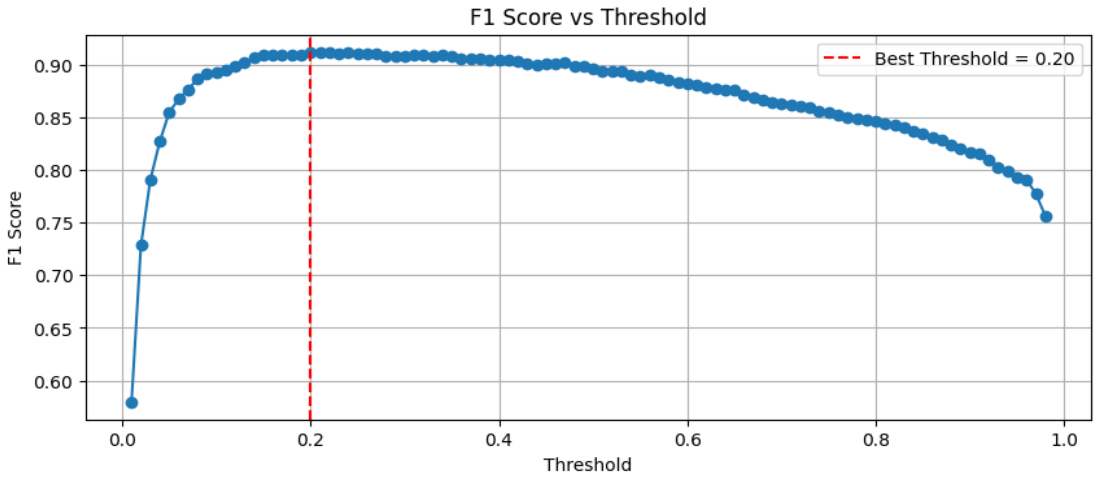
The LSTM autoencoder demonstrated strong capability for capturing the sequential structure of normal transaction behaviour and for identifying deviations through reconstruction error. Transactions with higher reconstruction errors were inferred to deviate from learned normal patterns and were consequently flagged as anomalous. These anomaly scores, combined with deep temporal features, were used as inputs for the subsequent supervised learning stage using an ANN. This approach provided a robust foundation for hybrid anomaly detection, enhancing the model's ability to effectively distinguish between normal and suspicious transaction patterns based on learned temporal behaviours.

Model Training Performance

During training, the ANN classifier demonstrated rapid convergence. Early stopping was applied based on validation loss, with a patience of 5 epochs, to prevent overfitting. The model consistently maintained validation accuracy above 99.88%, with a peak of 99.92% and a validation loss of 0.0046, indicating strong generalisation performance on unseen data. The plot of F1 score versus threshold in Figure 5 shows how the model’s classification performance varies with different decision thresholds. In the context of binary classification, the threshold is a critical parameter that determines the cutoff point for converting predicted probabilities into discrete class labels, specifically, whether a transaction is classified as suspicious or normal. For instance, when a model outputs a probability indicating the likelihood of fraud, a threshold (i.e., 0.5 by default) is applied: predictions above the threshold are labelled as suspicious, while those below are considered normal. Adjusting the threshold controls the model's sensitivity (i.e., recall) and specificity (i.e., precision). Lowering the threshold increases recall (detecting more true fraud cases) but may reduce precision (more false alarms), whereas raising it improves precision but risks missing actual fraudulent activity. Therefore, selecting an appropriate threshold is essential for balancing these competing objectives. As shown in the figure, tuning the threshold directly influences this balance, making it a vital step in optimising model performance for real-world fraud detection scenarios.

Figure 5

F1 Score vs. Threshold



As shown in Figure 5, the F1 score increases sharply at lower thresholds, peaks, and then gradually declines as the threshold becomes more conservative. The optimal threshold was empirically determined to be 0.20, where the F1 score reaches its maximum. This value offers the best balance between precision (reducing false positives) and recall (capturing true positives), which is critical in imbalanced datasets such as financial anomaly detection. The choice of F1 score as the primary evaluation metric is motivated by its ability to combine precision and recall into a harmonic mean, making it well-suited for scenarios where both false positives and false negatives carry significant consequences. Thus, this analysis underscores the importance of threshold tuning in achieving optimal model performance in real-world fraud detection tasks.

The proposed hybrid model demonstrates strong performance in classifying normal and anomalous financial transactions. Out of all actual normal transactions, 199,969 were correctly identified, with only 54 instances misclassified as anomalies (i.e., false positives). For anomalous cases, 864 transactions were correctly detected, while 113 were incorrectly predicted as normal (i.e., false negatives). These results highlight the model's high overall accuracy and its effectiveness at identifying true anomalies, with a relatively low false-positive rate. The robustness of this performance is attributed to the synergy of three components: the sliding window segmentation that captures short-term behavioural patterns, the LSTM-autoencoder that learns normal temporal dynamics and identifies deviations through reconstruction error, and the ANN that classifies these patterns using hybrid deep features.

However, despite its strong metrics, a closer examination reveals opportunities to improve handling of minority-class anomalies. The 113 false negatives, anomalies misclassified as normal, highlight the challenges of detecting subtle fraudulent behaviour in highly imbalanced datasets, where normal transactions vastly outnumber anomalies. In these scenarios, models tend to be biased toward the dominant class, which can hinder sensitivity to rare yet critical events. Additionally, the feature overlap between legitimate and fraudulent patterns may lead to ambiguity, particularly if some anomalies resemble normal behaviour in terms of sequence dynamics. These findings highlight the importance of improving the model's ability to distinguish between normal and anomalous transactions by using more informative features and applying techniques to address class imbalance, thereby ensuring more reliable detection of rare but impactful anomalies in financial systems.

Overall, the performance metrics in Table 2 demonstrate the robustness and effectiveness of the proposed hybrid model, which integrates sliding-window segmentation, an LSTM-autoencoder, and ANN classification.

Table 2

Performance Metrics of the Proposed Hybrid Model

Evaluation Metrics	Value
Accuracy	99.92%
Precision	94.12%
Recall	88.43%
F1 Score	91.19%

The high accuracy value highlights the model's overall reliability, while strong precision and recall indicate its ability to detect anomalies with minimal false alarms and missed detections. The balanced F1 score further confirms that the model maintains consistency between sensitivity and reliability in fraud detection scenarios. This robustness is achieved through the synergy of three components: the sliding window captures temporal patterns over fixed intervals, the LSTM-autoencoder learns normal behaviour and quantifies deviations through reconstruction error, and the ANN classifies sequences based on both deep temporal features and anomaly scores. Together, this hybrid architecture allows the model to generalise effectively across a large transaction dataset and consistently identify suspicious behaviour, even when anomalies are rare and subtle.

CONCLUSION

This study proposed a robust hybrid machine learning model for anomaly detection in banking transactions, integrating sliding-window segmentation, an LSTM-autoencoder, and an ANN classifier. The use of a 7-day sliding window enabled the model to capture temporal patterns and behavioural trends, which are often missed in static transaction analysis. The LSTM-autoencoder, trained exclusively on normal sequences, successfully learned the structure of typical transaction behaviour and identified deviations through reconstruction errors. These errors, combined with the encoded latent features, formed a comprehensive hybrid feature set that served as input to the ANN classifier.

The model demonstrated strong performance with an accuracy of 99.92%, precision of 94.12%, recall of 88.43%, and an F1 score of 91.19%, indicating a balance between detection sensitivity and reliability, with recall ensuring that most fraudulent transactions are successfully identified. In contrast, precision minimises false alarms by ensuring that the majority of flagged cases are indeed anomalies. This balance, reflected in the F1 score, demonstrates the model’s capability to provide both effective fraud detection and dependable predictions in real-world financial scenarios. The integration of these three components, a sliding window for temporal segmentation, an LSTM-autoencoder for unsupervised anomaly scoring, and an ANN for supervised classification, has proven to be a powerful and scalable approach. Overall, the hybrid model demonstrates robustness in handling complex, imbalanced financial data, offering a promising solution for real-time, adaptive fraud detection systems. By combining sliding-window segmentation, an LSTM-autoencoder, and an ANN within a layered framework, the proposed approach not only enhances anomaly detection performance but also improves the model’s generalizability in streaming transaction environments.

For future work, several directions may be explored to enhance the model's applicability and performance further. First, to better address class imbalance, future work could explore integrating advanced data-level techniques, such as Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), or generative models like GANs, to synthetically balance the training data and improve sensitivity to rare fraudulent cases. Additionally, cost-sensitive learning or class-weighted loss functions could be incorporated into the ANN training process to penalise misclassification of anomalies more heavily. Second, future studies could explore the use of attention-based architectures, particularly Transformer models, which are well-suited for capturing complex temporal dependencies and highlighting the most relevant parts of a transaction sequence. By leveraging their ability to model long-range relationships and subtle behavioural shifts, Transformers may improve the model's ability to distinguish between normal and anomalous patterns, especially in cases involving staged, delayed, or low-signal fraudulent behaviour. Third, real-time deployment on streaming platforms such as Apache Kafka or Flink could be investigated to evaluate the model's performance in production-scale environments. Finally, the incorporation of explainability techniques such as SHapley Additive exPlanations (SHAP) or local interpretable model-agnostic explanations (LIME) would provide greater transparency into model decisions, which is essential for operational deployment in regulated financial systems.

ACKNOWLEDGMENT

This research was partially funded by the Department Research Grants and Scientific Research Grants from Institut Teknologi Sepuluh Nopember (ITS) 2025, and the Fundamental Research Grants BIMA from the Ministry of Higher Education.

REFERENCES

Alarfaj, F. K., & Shahzadi, S. (2025). Enhancing fraud detection in banking with deep learning: Graph neural networks and autoencoders for real-time credit card fraud prevention. *IEEE Access*, *13*, 20633–20646. https://doi.org/10.1109/access.2024.3466288

Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *IEEE Access*, *11*, 137188–137203. https://doi.org/10.1109/access.2023.3339226

Anggraini, R. N. E., Ainurrochman, & Sarno, R. (2023). *Anomaly detection in raw audio using extreme learning machine.* 2023 14th International Conference on Information & Communication Technology and System (ICTS) (pp. 238–242). IEEE. https://doi.org/10.1109/icts58770.2023.10330852

Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection. *IEEE Access*, *12*, 64551–64560. https://doi.org/10.1109/access.2024.3394528

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2023). Credit card fraud detection in the era of disruptive technologies: A systematic review. *Journal of King Saud University - Computer and Information Sciences*, *35*(1), 145–174. https://doi.org/10.1016/j.jksuci.2022.11.008

Dastidar, K. G., Caelen, O., & Granitzer, M. (2024). Machine learning methods for credit card fraud detection: A survey. *IEEE Access*, *12*, 158939–158965. https://doi.org/10.1109/access.2024.3487298

Faisal, N. A., Nahar, J., Sultana, N., & Mintoo, A. A. (2024). Fraud detection in banking leveraging AI to identify and prevent fraudulent activities in real-time. *Journal of Machine Learning, Data Engineering and Data Science*, *1*(01), 181–197. https://doi.org/10.70008/jmldeds.v1i01.53

Farheen & Kumar, R. (2025). Sliding window-based anomaly detection. *Procedia Computer Science, 258,* 2520–2529. https://doi.org/10.1016/j.procs.2025.04.514

Ghimire, A. (2025). AI-Powered anomaly detection for AML compliance in US banking: Enhancing accuracy and reducing false positives. *Global Trends in Science and Technology, 1*(1), 95–120. https://doi.org/10.70445/gtst.1.1.2025.95-120

Hilal, W., Gadsden, A. S., & Yawney, J. (2021). A review of anomaly detection techniques and applications in financial fraud. *Expert Systems with Applications*, *193*(1), 116429. https://doi.org/10.1016/j.eswa.2021.116429

Ileberi, E., & Sun, Y. (2024). A hybrid deep learning ensemble model for credit card fraud detection. *IEEE Access, 12,* 175829–175838. https://doi.org/10.1109/access.2024.3502542

Iqbal, A., & Amin, R. (2024). Time series forecasting and anomaly detection using deep learning. *Computers & Chemical Engineering*, *182*, 108560. https://doi.org/10.1016/j.compchemeng.2023.108560

Irshad, F., Alkhalifah, T., Alturise, F., & Daanial Khan, Y. (2024). GCF-MLD: Integrated approach for money laundering detection using machine learning and graph network analysis. *IEEE Access, 12,* 183961–183972. https://doi.org/10.1109/access.2024.3510115

Ketenci, U. G., Kurt, T., Onal, S., Erbil, C., Akturkoglu, S., & Ilhan, H. S. (2021). A time-frequency based suspicious activity detection for anti-money laundering. *IEEE Access, 9,* 59957–59967. https://doi.org/10.1109/access.2021.3072114

Koo, K., Park, M., & Yoon, B. (2024). A suspicious financial transaction detection model using autoencoder and risk-based approach. *IEEE Access, 12,* 68926–68939. https://doi.org/10.1109/access.2024.3399824

Kute, D. V., Pradhan, B., Shukla, N., & Alamri, A. (2021). Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–A critical review. *IEEE Access, 9,* 82300–82317. https://doi.org/10.1109/access.2021.3086230

Paldino et al. (2024). The role of diversity and ensemble learning in credit card fraud detection. *Advances in Data Analysis and Classification*, *18,* 193–217. https://doi.org/10.1007/s11634-022-00515-5

Sehrawat, D., & Singh, Y. (2023). Auto-Encoder and LSTM-based credit card fraud detection. *SN Computer Science*, *4*(5). https://doi.org/10.1007/s42979-023-01977-w

Wardhani, T. P. M., Tahir, Z., Warni, E., Bustamin, A., Imran Oemar, M. A. F., & Kayyum, M. A. (2025). Deep learning approach in seismology: Enhancing earthquake forecasting using K-means clustering and LSTM networks. *Journal of Information and Communication Technology*, *24*(1), 29-51. https://doi.org/10.32890/jict2025.24.1.2

Yu, Q., Xu, Z., & Ke, Z. (2024*). Deep learning for cross-border transaction anomaly detection in anti-money laundering systems.* 2024 6th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) (pp. 244–248). IEEE. https://doi.org/10.1109/mlbdbi63974.2024.10823769

Zhang, Z., Chen, L., Liu, Q., & Wang, P. (2020). A fraud detection method for low-frequency transaction. *IEEE Access, 8,* 25210–25220. https://doi.org/10.1109/access.2020.2970614