Volume 12, No.9, September 2025

Journal of Global Research in Mathematical Archives

ISSN 2320 – 5822

RESEARCH PAPER

Available online at http://www.jgrma.com

IMPROVING TELECOMMUNICATION (TELECOM) FRAUD DETECTION ACCURACY THROUGH ADVANCED ARTIFICIAL INTELLIGENCE MODELS

Dr. Jvalantkumar Kanaiyalal Patel[1]

[1] Assistant Professor, Shri Manilal Kadakia College of Commerce, Management, Science and Computer Studies, Ankleswar
jvalant007@gmail.comcom

Abstract: The criminal justice system addresses telecom fraud as a major concern. Artificial intelligence development has led to increasingly sophisticated and creative telecom fraud texts. Unfortunately, when it comes to detecting telecom fraud in real-time, current security techniques like cell number tracking and detection and conventional machine-learning-based text identification aren't exactly top choices. An LSTM network-based deep learning architecture is suggested for the aim of identifying fraudulent behavior in CDRs in this study. The preprocessing involved in the methodology is extensive, including data cleaning, elimination of irrelevant attributes, critical discovery of outliers, and random under-sampling to curb extreme cases of class imbalances. A robust set of input variables was constructed using feature extraction and engineering methods and performance estimation was performed using the stratified k-fold cross-validation technique. Temporal relationships and complex patterns in sequential telecom data were used to train the LSTM model. Its performance was evaluated in comparison to NN, QA, and LR, three more strategy-based models. With an F1-score of 99.53, a recall of 99.37, a precision of 99.68, and an area under the curve (AUC) of 1.0000, the results demonstrate that the suggested LSTM model performed better than the competition. Instead, NN, QA and LR provided lower and less consistent results. The results validate the strengths, elasticity, and real-time applicability of the presented approach in detecting telecom fraud on large scale.

Keywords: Telecommunication Fraud Detection, Call Detail Records (CDRs), SIM-Box Fraud, Machine Learning and Deep Learning.

1. INTRODUCTION

Telecommunication fraud has emerged as a significant and evolving threat in the modern digital era, causing substantial financial losses and eroding customer trust [1][2]. Severe societal risks and substantial victim losses are a result of this new type of non-contact crime, which uses the Internet and telecommunication platforms like telephone, SMS, WeChat, QQ, and others to instantly and anonymously target large populations [3][4]. The rapid growth of the Internet has increased efficiency but has also opened doors to criminal activity and given con artists a larger target to exploit.

Voice conversations, text messages, Internet usage, and consumer activity create enormous amounts of data for the telecommunications business, which is one of the most data-intensive sectors [5][6][7]. This explosion of data presents both opportunities for service optimization and challenges in ensuring operational security [8]. Telecom systems produce diverse datasets, such as customer usage patterns, network traffic records, and operational metrics, all of which must be efficiently stored, processed, and analyzed to maintain system performance, customer satisfaction, and competitive advantage [9][10]. But with the advent of telecommunications networks that have the legacy systems (e.g., the Electronic Worldwide Digital Switch (EWSD) giving way to sophisticated IP-based revenue-generating or communication-carrying systems (e.g., Voice over Internet Protocol (VoIP) system, fraud detection has become very complex [11]. Although VoIP is more flexible, and efficient, it presents new vulnerabilities, and as a result, new fraud tactics are applied.

The concept of big data in a world where fraud is executed in a dynamic and adaptive manner, traditional rule-based fraud detection systems with their use of fixed thresholds and pre-determined conditions are simply becoming insufficient [12]. A new paradigm has emerged, centred around AI-based methods such as ML and DL, which can learn from both historical and real-time data, spot anomalies, and adjust patterns to accommodate novel fraud scenarios [13][14]. The ML and DL algorithms have proven to deal with large-scale datasets like Call Detail Records (CDRs) in order to identify complex fraudulent behaviours accurately and in real-time. In addition, deep learning models are self-learning and thus, continually refined to achieve a higher detection rate as they learns over time about changes in network traffic patterns and threat environment [15][16][17]. Telecom fraud detection using AI not only helps increase operational efficiency and service reliability, but also improves security considerably. Incorporating the power of advanced AI models, telecom operators able to reach a new level of detection performance, reduce financial losses, and retain customer trust in an ever more complex and hostile digital world [18]

1.1 Significance and Contribution of Study

Telecommunication fraud is a growing and more serious issue that uses the new medium of communication like VoIP, SMS and social media messaging to carry out mass, anonymous and high-impact attacks. The time-dependent and fast-changing nature of these

fraudulent schemes, coupled with the extensive amount of telecom data and complexity, meant that the traditional techniques of rule-based detection are not adequate. The legacy methods are not flexible when it comes to changing fraud patterns, which leads to a late discovery, loss of money and customer confidence. The relevance of this issue consists of a direct influence on operational safety, reliability of telecom services, and financial sustainability of telecom operators. Intelligent, adaptive and scalable intelligence solutions are needed to address this challenge, with processing high dimensions, sequential data available real-time such as Call Detail Records (CDRs), at high accuracy of detection and low false positive rates. In what follows, this work makes four main contributions:

- Developed a robust deep learning-based system capable of detecting SIM-box and other telecom frauds in real time using large-scale Call Detail Records (CDRs).
- Developed an effective preprocessing pipeline including data cleaning, irrelevant attribute removal, outlier detection, and class balancing via random under-sampling to improve data quality and model robustness.
- Presented a detection technique for time-dependent patterns and dependencies in CDR data using LSTM networks and deep learning.
- The F1-score, recall, precision, and accuracy of identification were all quite high.

1.2 Justification and Novelty

The novelty of this study is that it combines an LSTM-based deep learning architecture with an elaborate preprocessing pipeline to detect telecommunication fraud based on Call Detail Records (CDRs). By contrast to the conventional machine learning methods using fixed features, as the proposed method exploits the capacity of LSTM to trace the temporal sequencing relationships and sequential patterns in telecom data, the detection much more accurate. Some sophisticated preprocessing which includes outlier management, irrelevant attributes elimination, and class balancing through random under-sampling shall help to maintain the data quality and increase the quality of the model. Comparative evaluation against Neural Network (NN), Qwen2-Audio (QA), and Logistic Regression (LR) models demonstrates that the proposed LSTM significantly outperforms these baselines, achieving near-perfect classification metrics with an AUC of 1.0000. This performance, combined with the model's scalability and adaptability, justifies its suitability for real-time deployment in large-scale telecom environments where rapid, accurate, and reliable fraud detection is critical.

1.3 Structure of Paper

The following is the outline of the paper: Section 2 discusses previous research on telecom fraud detection, Section 3 describes the technique, Section 4 gives the results and analysis, and finally, Section 5 summarizes the main points and discusses potential future directions.

2. LITERATURE REVIEW

This section presents research on Telecommunication (Telecom) Fraud Detection utilizes diverse Artificial Intelligence techniques; the summary of these studies is provided in Table 1.

Jiang et al. (2025) reduce the amount of background noise caused by irrelevant neurons, making the model more accurate and efficient to run. The LENS-RMHR model enhances feature representation capacities and training efficiency through the use of residual connections, a multi-head attention mechanism known as RoBERTa. Their enlarged dataset, which includes eight separate categories covering a wide range of fraud forms, is based on the CCL2023 telecoms fraud dataset. Also, in multi-classification scenarios, the model's discovery has been improved by using a dual-loss function. The observed experimental results show that LENS-RMHR performs better when applied to various benchmark datasets, which bodes well for its future use in text categorization and telecom fraud detection [19].

Mishra and Shivaji (2025) that anomaly detection, decision trees and clustering are among the machine learning algorithms used in the case in mapping unusual call patterns, unauthorized access and Identity theft. In order to detect anomalous behavior in time domain data and frequency domain, training and testing models are used on time-domain and frequency domain data of various telecom networks. Improving the accuracy and efficiency of fraud detection networks is the goal of this project, which draws on prior work in DL, NB, and SVM. The accuracy in the classification and prediction of fraudulent actions of these models is used as the measure of performance, thus decreasing the likelihood of false negative/false positive predictions [20].

Zhang and Jiang et al. (2024) The RoBERTa-MHARC model for text-based telecom fraud detection is shown here. RoBERTa is made up of a multi-head attention mechanism and residual links. This model constructs a five-category dataset that includes customer service impersonation, leadership acquaintance impersonation, loans, public security fraud, and normal text by combining basic samples from the CCL2023 telecom fraud dataset with the gathered telecom fraud text data. The model achieves better training efficiency by utilizing residual connections and a multi-head attention function. The model enhances its accuracy in multi-class classification by merging the cross-entropy loss with the inconsistency loss function. Impressive F1 scores on the FBS dataset (97.65), the own dataset (98.10), and the news dataset (93.69) demonstrate that the model performs admirably when trained on multiple benchmark datasets [21].

Li et al. (2023) employ attribute dependencies as they relate to the entropy function to improve data quality, which in turn fixes the problem of telecom fraud detection using variance data. Their max-correlation and max-independence rate (MCIR) method of attribute reduction aims to improve data quality by removing points that are redundant or noisy. Next, in order to deal with incomplete or missing data, they suggest MCIR-RGAD, which stands for maximum consistent block-based rough-gain anomaly detection. Last

but not least, results from testing on actual UCI and telephonic fraud data demonstrated that the MCIR-RGAD approach effectively reduced computing time, improved data quality, and calculated incompleteness data [22].

Hong, Connie and Ong Goh, (2023) suggest an innovative approach to detecting and classifying scam calls through the use of DL and NLP through machine learning. A dataset of both valid and unscammed chats can be used to train a model that can learn and understand the caller's circumstances and determine if the chat is a scam or not. To build a reliable and accurate classifier, they use natural language processing techniques including word embeddings, preprocessing, and audio sample to text conversion with the Google API. The LSTM algorithm outperformed the others, providing 85.61 percent accurate scam call detection [23].

Mohana et al. (2022) explore many methods for identifying and avoiding fraud in the field of communications. Various telecom fraud classifications are discussed in this study, along with detection challenges and possible remedies. After you've outlined the current methods' effectiveness, you should provide them some pointers on how to do better and a framework for selecting the right metrics to monitor their performance [24].

Acevedo-Viloria et al. (2021) the effectiveness of integrating consumer data from super-apps, data from mobile phone lines, and traditional credit risk indicators at the feature level in detecting credit card fraud involving identity theft sooner. A ROC AUC score of 0.81 was achieved, indicating increased performance, by using a model that takes in both traditional and alternative credit bureau data as inputs, thanks to the proposed methodology. The technique is tested using a credit lender's digital platform database, which contains information about 90,000 users. For this assessment, they accounted for both traditional ML metrics and monetary costs [25]

Table 1: Comparative Analysis of Machine Learning Techniques for Telecommunication (Telecom) Fraud Detection

Author(s) & Year	Methodology	Data	Key Findings	Limitations	Future Work
Jiang et al. (2025)	The LENS-RMHR model employs a dual-loss function, residual connections, multi-head attention, and RoBERTa for multi-class classification.	CCL2023 telecom fraud dataset (expanded to 8 categories)	Effectively mitigates noise from irrelevant neurons; improved accuracy and computational efficiency; strong performance across multiple datasets	Scalability to real-time fraud detection not tested	Extend to real-time detection and cross-lingual telecom fraud scenarios
Mishra & Shivaji (2025)	ML algorithms (clustering, decision trees, anomaly detection, Naïve Bayes, SVM, deep learning) on time and frequency domain features	Telecom network data with unusual call patterns, unauthorized access, identity theft cases	Accurate classification and prediction of fraudulent activities; reduced false negatives and positives	Lack of evaluation on large-scale streaming data	Integrate with real-time anomaly detection pipelines
Zhang & Jiang et al. (2024)	A multi-head attention and residual connection cross-entropy loss function that is inconsistent Reach out to RORETA-MHARC	Modified CCL2023 dataset with 5 fraud categories	F1 scores were 97.65 on the FBS dataset, 98.10 on the own dataset, and 93.69 on the news dataset.	Limited to textual fraud data	Expand to multimodal fraud detection (text + call metadata)
Li et al. (2023)	Attribute reduction using MCIR, anomaly detection via MCIR-RGAD	Telecom fraud data & UCI dataset	Improved data quality, reduced computation time, handled incomplete data	Performance on high-dimensional big data not explored	Apply to large-scale heterogeneous telecom data sources
Hong, Connie & Ong Goh (2023)	NLP and deep learning (LSTM) with Google API for audio-to-text conversion and scam call detection	Scam & non-scam call dataset	LSTM achieved 85.61% accuracy in detecting scam calls	Accuracy relatively low compared to state-of-the-art	Explore transformer-based architectures for improved detection
Mohana et al. (20220	Literature review on telecom fraud types, detection issues, and solutions	Multiple literature sources	Provides classification, challenges, and metric selection guidance	No experimental validation	Conduct empirical studies to evaluate proposed solutions
Acevedo-Viloria et al. (2021)	Feature-level merging of data from super-apps, mobile phones, and credit risk factors	~90,000 user records from a lender’s digital platform	Improved identity theft detection; ROC AUC = 0.81	Focused only on credit card fraud	Expand framework to cover broader telecom fraud scenarios

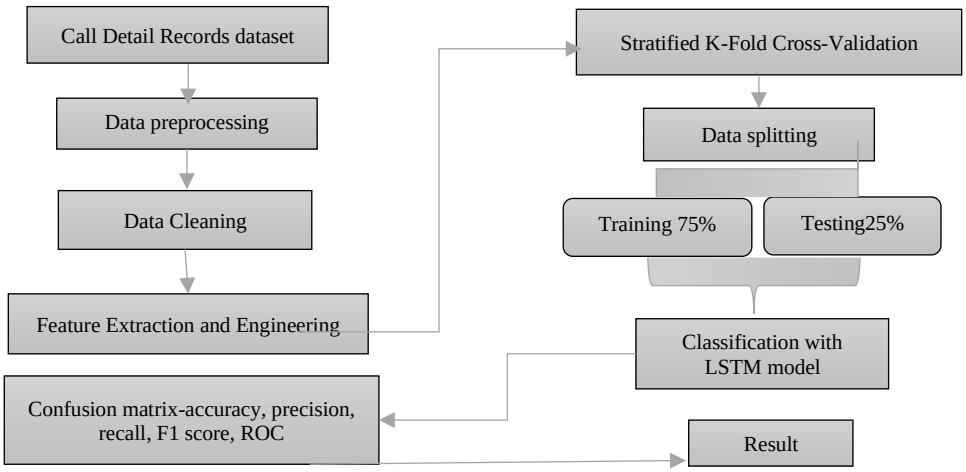

Figure 1: Flowchart of the Telecommunication (Telecom) Fraud Detection

3. METHODOLOGY

The methodology for telecommunication fraud detection, as illustrated in Figure. 1, begins with the acquisition and initial pre-processing of the Call Detail Records (CDR) dataset. In the pre-processing phase, data is cleaned, missing values are dealt with, and irrelevant traits are filtered out. This is followed by feature extraction and engineering to create a full set of input variables for training the model. Training and testing sets are created from the dataset using stratified k-fold cross-validation, which maintains the same class distribution. Next, an LSTM network is given a classification job. This network has been trained to find and account for temporal dependencies in sequential telecom data. A model's success is judged by its accuracy, precision, recall, and F1-score, among other things. The Receiver Operating Characteristic (ROC) graph can help you figure 1 out how well the classification works.

The following sections provide each step description which also shows in methodology and proposed flowchart:

3.1 Data Collection

In telecom fraud detection, acquiring and preparing Call Detail Records (CDRs) is foundational. CDRs typically include metadata such as calling and called numbers, timestamps, duration, cell tower location, and call type. These records are aggregated from multiple network sources switches, routers, billing systems to assemble a comprehensive dataset.

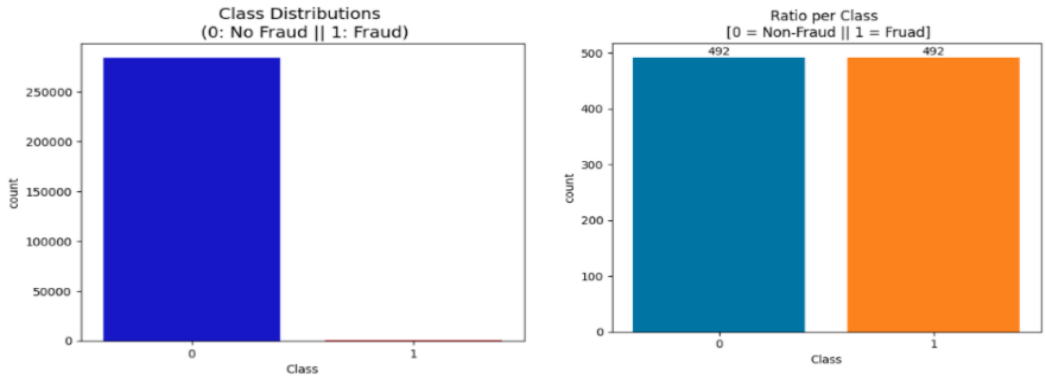

Figure 2: Target Class Distribution: (Left) (Imbalanced), (Right) (Balanced)

Figure 2 shows both the unbalanced dataset and the balanced dataset that was achieved by randomly under sampling some cells. For large datasets with significant imbalances, random under-sampling is a common technique. Methodically removing samples at random from the majority class is what it comprises. Less than one percent of transactions were fraudulent, demonstrating a clear class imbalance in the initial data.

3.2 Data Pre-processing

Effective detection of SIM box fraud requires clear, relevant data, which must be pre-processed. It is necessary to remove duplicates, anomalies, and missing information from the raw data, which is usually Call Detail Records (CDRs). Look for and analyse possible warning signs, often called outliers. Feature engineering, which involves making changes to current features in order to build new ones, is the next stage. What follows is a discussion of the pre-processing procedures:

- Removing duplicates and non-relevant call types (e.g., fixed-line or internal administrative calls.
- Dropping records with critical missing fields if imputation would introduce bias or poor model performance.

- Outlier Detection Using statistical methods like Z-scores or IQR to detect anomalies (e.g., extremely long call durations), followed by capping or transformation to avoid skewed distributions.
- In order to ensure that no values are missing or incorrect, they check all call detail records. So, for instance, if the length of the calling number is under 12 characters, the record will be discarded and fixed appropriately.

3.3 Feature Extraction and Engineering

Extracting useful characteristics was the first step in getting the data ready for the ML model. Here, characteristics for every SIM card were retrieved [26]. In order to train the model with the finest possible features, features were also engineered. In order to create more valuable features, feature engineering employs techniques like feature selection and dimensionality reduction.

3.4 Stratified K-Fold Cross-Validation

The Elbow Method is employed to ascertain the K value in this investigation. By comparing the % of results included inside each cluster size to the number of clusters that form an elbow at a particular position, you can quickly find the best number of clusters using this technique [27]. Using machine learning, the author of this piece on K-means clustering found the Elbow Method, which counts the clusters in the dataset.

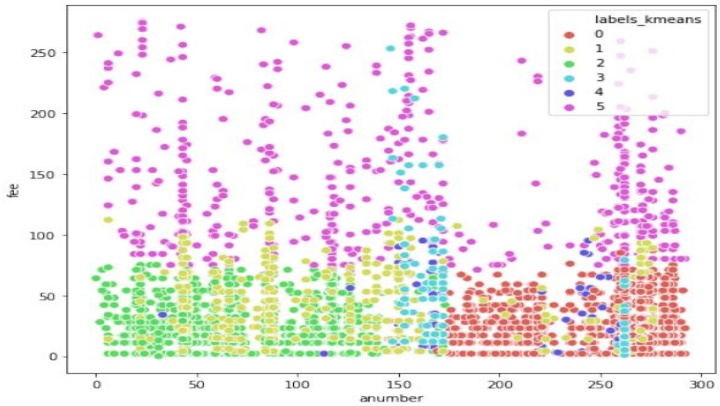

Figure 3: K-Means Clustering

The Elbow method's curve value is displayed in Figure 3. It is clear from the image that six clusters are the upper limit.

3.5 Data Splitting

The dataset splitting was done so that the model could be used in more situations. The ratio used for the splitting was 75:25. Model testing assessments made use of 25% of the dataset, while the proposed model training used 75% of the dataset.

3.6 Classification with Long Short-Term Memory (LSTM)

This study employs an enhanced LSTM model to circumvent the caveats of conventional RNNs, particularly with regard to the vanishing and exploding gradient difficulties. LSTM neural networks differ from the more basic RNNs in that they use a four-layer interacting repeating module [28]. Modules like this are perfect for time series jobs because they enable the model to remember key data over extended sequential times. Pictured in Figure 4 is the architecture of the LSTM cell.

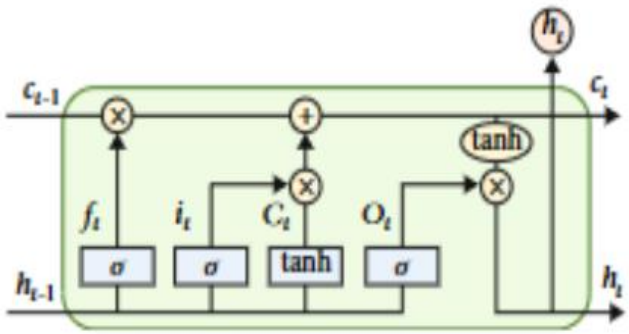

Figure 4: The LSTM Cell Structure

At the core of each LSTM unit are three gates:

Forget Gate: Determines the amount of historical data that is eliminated when utilizing the sigmoid activation; it is obtained from Equation (1):

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

Input Gate: Uses two equations to determine what new data to include Equation (2):

$$i_t - \sigma(W_i.[h_{t-1}, x_t] + b_i), \tilde{C}_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \quad (2)$$

Cell State Update: Integrates previously stored information with newly proposed values as seen in Equation (3):

$$C_t = f_t.C_{t-1} + i_t.\tilde{C}_t \quad (3)$$

Output Gate and Hidden State: Finding the end output at each timestep is done using Equation (4):

$$o_t - \sigma(W_o.[h_{t-1}, b_o] + b_o), f_t = o_t.tanh(C_t) \quad (4)$$

The model's ability to efficiently filter, update, and output time-dependent characteristics is made possible by these gated mechanisms, which in turn facilitate accurate forecasting and robust long-term learning. LSTM is better than regular RNNs at spotting relationships and patterns in sequential data.

3.7 Evaluation Methods

A classification system's performance can be better understood with the help of the confusion matrix. To further understand the model's predictions, it provides a matrix containing the counts of TP, TN, FP, and FN. Model recall, precision, and accuracy can be easily understood with the help of this matrix.

3.7.1. Accuracy

One simple statistic is accuracy, which is just the percentage of cases (fraudulent and non-fraudulent transactions included) that were correctly classified relative to the total number of instances. In Equation (5):

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

3.7.2. Precision

A model's accuracy in identifying instances of fraud relative to the total number of cases categorised as such is known as its precision, or Positive Predictive Value (PPV) in Equation (6):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

3.7.3. Recall

Equation (7) provides the formula for calculating recall, which is sometimes called sensitivity or the TPR. It is a measure of the model's accuracy in identifying fraud instances:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

3.7.4. F1_Score

A single metric that harmonically balances Precision and Recall is provided by the F1 Score Equation (8):

$$F1 - score = 2 * \frac{(precision*recall}{(precision+recall)} \quad (8)$$

3.7.5. ROC

The evaluation metrics indicated earlier were used in conjunction with one another to determine the final model. To address the dataset's imbalance, F1 Score and ROC-AUC were given special attention. Robust and scalable fraud protection was ensured by recommending the selected model for deployment in telecom banking's real-time fraud detection systems.

4. RESULTS AND DISCUSSION

ML and DL models were used to get these outcomes, and experiments were run using a Windows-powered, 7th-generation Intel Core i7 computer. Set up the Jupyter notebook and Python 3.x. The proposed LSTM model's performance test results on the Call Detail Records (CDR) corpus are shown in Table 2, demonstrating its usefulness for classification tasks. The model showed an accuracy of 99.76% that reflects the set of reliable general predictions, the precision of 99.68% proves successful in identifying positive values with low false positive rate. The recall score of 99.37% speaks to the high ability of the model in identifying true positive cases with the F1-score of 99.53% confirming a good trade-off between precision and recall. These findings confirm that LSTM model used in this work is resilient and effective in performing and categorizing CDR data with virtually perfect outcomes.

Table 2: Parameters' performance LSTM model on the Call Detail Records dataset

Measures	LSTM
Accuracy	99.76
Precision	99.68
Recall	99.37
F1-score	99.53

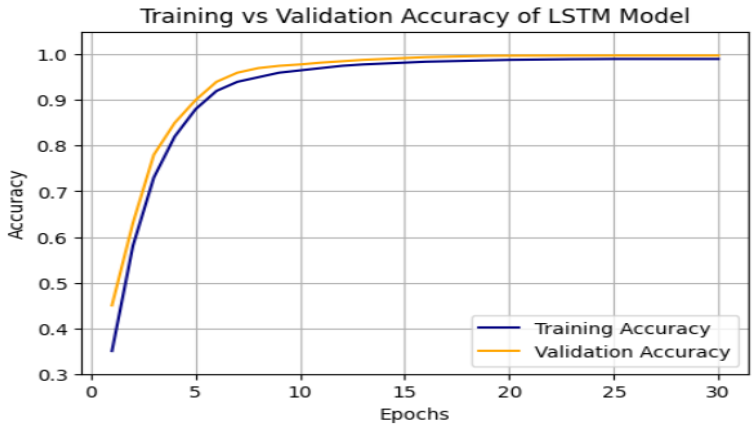

Figure 5: Accuracy Curve of LSTM

The accuracy of the proposed LSTM during training and validation across 30 epochs is shown in Figure 5. The accuracy, which can range from 0.3 to 1.0, is shown on the y-axis, while the total number of epochs is shown on the x-axis. Lines colored blue represent training accuracy and orange represent validation accuracy. The steps of both curves may be distinguished by a sharp increase in accuracy in the first epochs after which they gradually approach unity. Significantly, the validation accuracy is always slightly higher than the training accuracy during the majority of the training process, which implies the high capacity of generalization of the model and the lack of overfitting.

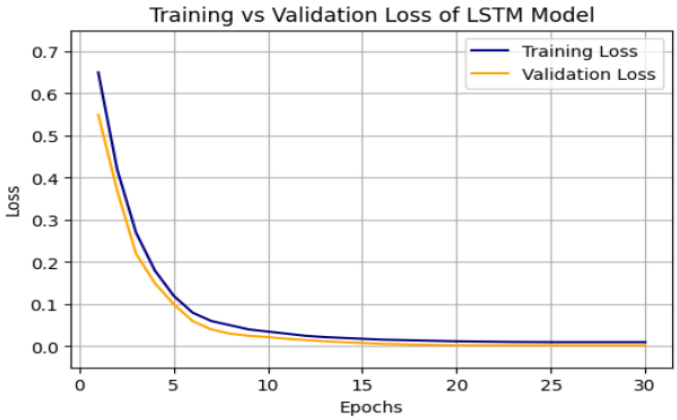

Figure 6: Training vs. Validation Loss of LSTM

Shown in Figure 6 are the training and validation loss curves, spanning 30 epochs, of the LSTM model suggested in this paper. The epoch number is shown on the first axis (x), and the recorded loss, which falls between 0.0 and 0.7, is shown on the second axis. The training loss is associated with the blue line, whereas the validation loss is represented by the orange line. The two curves show a steep decrease in the first epochs, which indicates the high rate of learning by the model and efficient minimization of the error. The curves cross beyond 10 15 epochs and start stabilizing at the values near zero, which indicates that the model managed to find the underlying patterns in the dataset and the generalization performance is high without substantial overfitting.

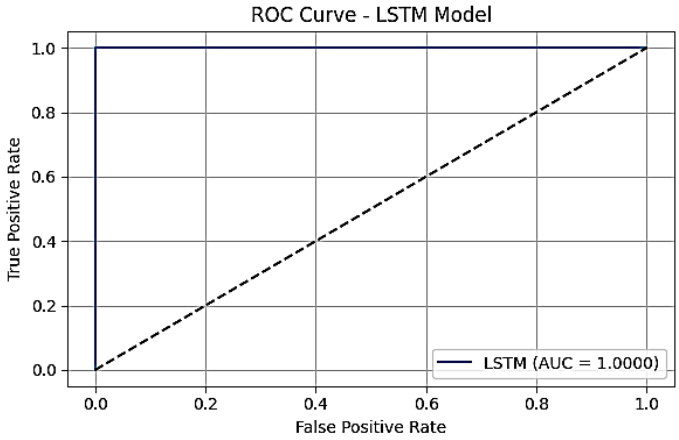

Figure 7: ROC of LSTM Model

Figure 7 shows the Receiver Operating Characteristic (ROC) of LSTM-based call fraud detection model. The model has the ability to categorise cases of fraud and normal with a hundred percent degree of accuracy, as seen by the curve, which demonstrates nearly flawless classification ability. The AUC value of 1.0000 was utilised to do this. The ROC curve's sharp rise to the top left corner indicates a high true positive rate and a low false positive rate.

Table 3: Performance Comparison of ML and DL models for Telecommunication (Telecom) Fraud Detection

Matrix	LSTM	NN[29]	QA [30]	LR[31]
Accuracy	99.76	83.67	61.83	95.86
Precision	99.68	72.60	68.40	96.4
Recall	99.37	92.17	51.12	94.8
F1-score	99.53	81.22	58.51	95.6

The outcomes of evaluating the different ML and DL models for identifying telecommunication fraud using several metrics, such as accuracy, precision, recall, and F1-score, are displayed in Table 3. With superior findings and an accuracy of 99.76%, the LSTM model demonstrates balanced and robust classification properties with high precision, recall, and F1-scores above 99. Comparatively, the Neural Network (NN) model registers a medium score of 83.67% accuracy, precision (72.60), F1-score (81.22), though it takes a relatively high value of recall (92.17). The QA (Qwen2-Audio) model has the lowest overall performance with the accuracy of 61.83% and huge disparity across the measures, such as the recall 51.12%. The Logistic Regression (LR) is relatively successful as compared to QA, achieving an accuracy of 95.86 and balanced precision, recalls as well as F1-scores of over 94%, but below the corresponding models with deep learning.

The LSTM model proposed to application in fraud detection in the telecommunication industry shows very high reliability and consistency on all the measuring-testing variables, which reveals that it is capable of detecting the fraud with an acceptable degree of accuracy, combined with minimal false positive and false negative outcomes. However, temporal sequence learning as a feature of LSTM models can help effectively define more complicated trends in the call data, allowing to detect them even in problematic settings. One style of performance that is well-known, good generalization to unseen data structures makes it suitable to be used in practical applications in telecom, where the scale of operation is large and many data structures are common to several applications.

5. CONCLUSION AND FUTURE SCOPE

The problem of telecommunication fraud has attracted the attention of service providers because it is dynamic and has far-reaching financial implications. The application introduced LSTM as a DL0 model into detecting fraudulent patterns in Call Detail Records (CDRs), taking an advantage of the DL model which can learn temporal correlation and complex patterns in sequential telecom data. This model performed quite well, with an F1-score of 99.53%, an accuracy of 99.76%, a precision of 99.68%, a recall of 99.37%, and an area under the receiver operating characteristic curve (ROC) of 1.0000. The multifactor analysis also proved the effectiveness and strength of the LSTM model because it is more powerful than other methods to be used in large-scale and real-time deployments in the telecom market.

The future work will be aimed at novel hybrid architectures where LSTM will be used concomitantly with attention mechanisms or with Transformer-based models to capture patterns. Learning methods cost-sensitive learning will be investigated as a means to improve class imbalance and optimization methods will be utilized to allow inference to execute with low latency in low-latency applications. The domain of fraud detection will further be extended to various types of fraud which will be verified at large datasets in order to be applicable and generalized to changing operational environment. “

REFERENCES

[1] P. Ni and Q. Wang, “Internet and Telecommunication Fraud Prevention Analysis based on Deep Learning,” *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2137630.

[2] H. Kali, "Optimizing Credit Card Fraud Transactions identification and classification in banking industry Using Machine Learning Algorithms," *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 11, pp. 85–96, 2024.

[3] V. Prajapati, "Enhancing Threat Intelligence and Cyber Defense through Big Data Analytics: A Review Study," *J. Glob. Res. Math. Arch.*, vol. 12, no. 4, 2025.

[4] V. Verma, "Deep Learning-Based Fraud Detection in Financial Transactions : A Case Study Using Real-Time Data Streams," vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[5] O. T. Odofin *et al.*, "Integrating Artificial Intelligence into Telecom Data Infrastructure for Anomaly Detection and Revenue Recovery," *ICONIC Res. Eng. JOURNALS*, vol. 5, no. 2, 2021.

[6] R. Patel, "Security Challenges In Industrial Communication Networks: A Survey On Ethernet/Ip, Controlnet, And Devicenet," *Int. J. Recent Technol. Sci. Manag.*, vol. 7, no. 8, 2022.

[7] J. Mishra, B. B. Biswal, and N. Padhy, "Machine Learning for Fraud Detection in Banking Cyber security Performance Evaluation of Classifiers and Their Real-Time Scalability," in *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, IEEE, Feb. 2025, pp. 431–436. doi: 10.1109/ESIC64052.2025.10962752.

[8] N. K. Prajapati, "Federated Learning for Privacy-Preserving Cybersecurity: A Review on Secure Threat Detection," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 4, pp. 520–528, Apr. 2025, doi: 10.48175/IJARSCT-25168.

[9] X. Hu, H. Chen, S. Liu, H. Jiang, G. Chu, and R. Li, "BTG: A Bridge to Graph machine learning in telecommunications fraud detection," *Futur. Gener. Comput. Syst.*, vol. 137, pp. 274–287, Dec. 2022, doi: 10.1016/j.future.2022.07.020.

[10] G. Mantha, "Transforming the Insurance Industry with Salesforce: Enhancing Customer Engagement and Operational Efficiency," *North Am. J. Eng. Res.*, vol. 5, no. 3, 2024.

[11] V. Shah, "Scalable data center networking : Evaluating virtual extensible local area network-Ethernet virtual private network as a next-generation overlay solution," *Asian J. Comput. Sci. Eng.*, vol. 8, no. 3, pp. 1–7, 2023.

[12] D. Patel, "Leveraging Blockchain and AI Framework for Enhancing Intrusion Prevention and Detection in Cybersecurity," *Tech. Int. J. Eng. Res.*, vol. 10, no. 6, pp. 853–858, 2023, doi: 10.56975/tijer.v10i6.158517.

[13] R. Patel, "Automated Threat Detection and Risk Mitigation for ICS (Industrial Control Systems) Employing Deep Learning in Cybersecurity Defence," *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, pp. 584–591, 2023, doi: 10.14741/ijcet/v.13.6.11.

[14] D. D. Rao, S. Madasu, S. R. Gunturu, C. D'britto, and J. Lopes, "Cybersecurity Threat Detection Using Machine Learning in Cloud-Based Environments: A Comprehensive Study," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 12, no. 1, 2024.

[15] E. Edozie, A. N. Shuaibu, B. O. Sadiq, and U. K. John, "Artificial intelligence advances in anomaly detection for telecom networks," *Artif. Intell. Rev.*, vol. 58, no. 4, 2025, doi: 10.1007/s10462-025-11108-x.

[16] D. Patel, "Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model," *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, pp. 576–583, 2023, doi: 10.14741/ijcet/v.13.6.10.

[17] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 10, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[18] D. D. Rao, A. A. Waoo, M. P. Singh, P. K. Pareek, S. Kamal, and S. V. Pandit, "Strategizing IoT Network Layer Security Through Advanced Intrusion Detection Systems and AI-Driven Threat Analysis," *J. Intell. Syst. Internet Things*, vol. 24, no. 2, pp. 195–207, 2024, doi: 10.54216/JISIoT.120215.

[19] L. Jiang *et al.*, "Telecom Fraud Recognition Based on Large Language Model Neuron Selection," *Mathematics*, vol. 13, no. 11, p. 1784, May 2025, doi: 10.3390/math13111784.

[20] N. Mishra and G. B. Shivaji, "Data Mining for Fraud Detection in Telecommunications: Detecting Anomalous Behaviors in Real-Time," in *2025 International Conference on Automation and Computation (AUTOCOM)*, IEEE, Mar. 2025, pp. 1340–1345. doi: 10.1109/AUTOCOM64127.2025.10957085.

[21] J. Li, C. Zhang, and L. Jiang, "Innovative Telecom Fraud Detection: A New Dataset and an Advanced Model with RoBERTa and Dual Loss Functions," *Appl. Sci.*, vol. 14, no. 24, pp. 1–16, 2024, doi: 10.3390/app142411628.

[22] R. Li, H. Chen, S. Liu, K. Wang, B. Wang, and X. Hu, "TFD-IIS-CRMCB: Telecom Fraud Detection for Incomplete Information Systems Based on Correlated Relation and Maximal Consistent Block," *Entropy*, vol. 25, no. 1, p. 112, Jan. 2023, doi: 10.3390/e25010112.

[23] B. Hong, T. Connie, and M. K. Ong Goh, "Scam Calls Detection Using Machine Learning Approaches," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, IEEE, Aug. 2023, pp. 442–447. doi: 10.1109/ICoICT58202.2023.10262695.

[24] R. H J and Mohana, "Fraud Detection and Management for Telecommunication Systems using Artificial Intelligence (AI)," in *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, Oct. 2022, pp. 1016–1022. doi: 10.1109/ICOSEC54921.2022.9951889.

[25] J. D. Acevedo-Viloria, S. S. Perez, J. Solano, D. Zarruk-Valencia, F. G. Paulin, and A. Correa-Bahnsen, "Feature-Level Fusion of Super-App and Telecommunication Alternative Data Sources for Credit Card Fraud Detection," in *2021 IEEE International Conference on Intelligence and Security Informatics (ISI)*, IEEE, Nov. 2021, pp. 1–6. doi: 10.1109/ISI53945.2021.9624796.

[26] I. Ighneiwa and H. S. Mohamed, "Bypass fraud detection: Artificial intelligence approach," *arXiv*, no. November 2017, pp. 3–6, 2017.

[27] M. Jabbar and S. Suharjito, "Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 5, no. 4, pp. 63–69, Jul. 2020, doi: 10.25046/aj050409.

[28] Y. Gao, D. Yin, X. Zhao, Y. Wang, and Y. Huang, "Prediction of Telecommunication Network Fraud Crime Based on Regression-LSTM Model," *Wirel. Commun. Mob. Comput.*, vol. 2022, pp. 1–16, Aug. 2022, doi: 10.1155/2022/3151563.

[29] Q. Zhao, K. Chen, T. Li, Y. Yang, and X. F. Wang, "Detecting telecommunication fraud by understanding the contents of a call," *Cybersecurity*, 2018, doi: 10.1186/s42400-018-0008-5.

[30] Z. Ma *et al.*, *TeleAntiFraud-28k: An Audio-Text Slow-Thinking Dataset for Telecom Fraud Detection*, vol. 1, no. 1. 2025.

[31] G. Liu, "Leveraging Machine Learning for Telecom Banking Card Fraud Detection: A Comparative Analysis of Logistic Regression, Random Forest, and XGBoost Models," *Comput. Artif. Intell.*, vol. 1, no. 1, pp. 13–27, 2024, doi: 10.70267/1cc7aw07.

