*International Journal of Engineering Research and Development*
*e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com*
*Volume 22, Issue 7 (July 2026), PP 45-52*

# A Comparative Analysis of Machine Learning Models for Fraud Detection in Telecommunication Billing Systems Using Simulated Call Detail Records

Uduak Joseph Essien
*Department of Electrical and Electronics Engineering Technology, Akwa Ibom State Polytechnic, Ikot Osurua, Nigeria*

Kufre Michael Udofia
*Department of Electrical and Electronics Engineering, University of Uyo, Uyo, Nigeria*

Emem Godwin Young
*Department of Computer Engineering Technology, Akwa Ibom State Polytechnic, Ikot Osurua, Nigeria*

Ayodele Anthony Olatunji
*Department of Electrical and Electronics Engineering, Michael Okpara University of Agriculture, Umudike, Nigeria*

***Abstract***
*This study presents a comparative analysis of machine learning models which are Logistic Regression, Random Forest, and Gradient Boosting for the detection and prevention of fraud in telecommunication billing systems. Fraudulent activities such as SIM box fraud, call masking, and subscription fraud threaten the financial sustainability of telecommunication network operators, leading to revenue leakage. A simulated Call Detail Record (CDR) dataset of 1000 billing records obtained from Kaggle was used. Data preprocessing, feature engineering, and random oversampling were applied to address class imbalance in the dataset. The models were evaluated using accuracy, precision, recall, F1-score, and the Area Under the ROC Curve (AUC). Before oversampling, all three models achieved high accuracy but zero recall for the fraud class, revealing their inability to detect fraudulent minority-class transactions. After random oversampling, Logistic Regression showed the most balanced improvement, with recall increasing from 9% to 48% and AUC improving from 0.46 to 0.73. The study concludes that while Logistic Regression is interpretable and suitable for real-time billing fraud prevention, the challenge of class imbalance remains a core limiting factor to model effectiveness, and future research should explore advanced resampling techniques and algorithm-level solutions.*
***Keywords:*** *Call Detail Records, Random Forest, Gradient Boosting, Logistic Regression, Oversampling, Fraud Detection, Telecommunication Billing Systems.*

---

Date of Submission: 02-07-2026 Date of acceptance: 11-07-2026

---

## I. INTRODUCTION

In recent years, telecommunication companies have been losing a lot of money because of fraud and fraudsters keep finding clever ways to cheat the system and get free or illegal access to services. This is happening more recently because telecommunication networks have become more complex because of the number of people using phones and internet services every day. Fraudsters use common tricks like Subscriber Identity Module (SIM) box fraud, call spoofing, and even hacking into customer accounts to steal their data or make unauthorized calls (Bello et al., 2023).

The systems that telecommunication companies used in the past to detect and mitigate these fraudulent activities worked by following simple rules like automatically flagging calls that lasted too long or came from suspicious locations. However, fraudsters are now employing new techniques. They keep changing their techniques and the old systems are outdated and increasingly inadequate in detecting and preventing advanced fraudulent activities. This is why telecommunication and other companies are now looking for more advanced ways, like using artificial intelligence, to detect and mitigate fraud in real time before the threats result in substantial losses (Hu et al., 2023).

Over the years, many companies have been using machine learning and deep learning especially the banking and finance companies where people sometimes try to cheat the system by making fake transactions or stealing money. These are smart computer programs that learn from data and they study millions of past transactions and understand what is normal so they can spot anything suspicious or unusual (Nguyen et al., 2020). However, directly applying these techniques to telecommunication billing fraud detection presents several challenges, including the need for real-time detection, handling continuous streams of billing data, and maintaining transparency in fraud decisions (Xu et al., 2023). Deep learning models are typically black-box systems, which makes it difficult for operators to understand why a particular transaction was flagged as fraudulent, which reduces trust and complicates regulatory compliance (Psychoula et al., 2021).

This study presents a comparative analysis of three supervised machine learning models which are Logistic Regression, Random Forest, and Gradient Boosting used for detecting and preventing fraud in telecommunication billing systems. The study employs a simulated Call Detail Record (CDR) dataset and addresses the common challenge of class imbalance through random oversampling. The study aims to identify the most suitable model for real-time, interpretable, and affordable fraud prevention in telecommunication billing environments.

### A. Statement of the Research Problem

The Telecommunication Billing systems are crucial to the telecommunication companies and the ability to precisely record, process, and bill customers for services. Traditional telecommunication billing systems, however, struggle with preventing fraud, particularly when it comes to identifying changing fraud patterns while preserving transparency and minimizing cost. The financial sustainability of telecommunication network operators is threatened by fraudulent activities such as SIM box fraud, call masking, and subscription fraud, which all lead to revenue leakage. According to Hu et al. (2021), existing rule-based fraud detection systems are static and unable to quickly adjust to new, more sophisticated fraud techniques. Meanwhile, more complex machine learning approaches frequently have high computational costs and limited explainability, making them impractical for many operators (Johnson et al., 2022).

### B. Objectives of the Study

The main objective of the study is to develop machine learning models to prevent telecommunication billing fraud by analyzing simulated call data in a controlled offline environment. The specific objectives are to: (i) use simulated telecommunication billing data to help prevent fraudulent activity by identifying patterns in call logs and other relevant features; (ii) identify and extract key features in the data that may signal fraudulent behavior within the billing system; (iii) develop user-friendly machine learning models using logistic regression, random forest and gradient boosting to distinguish between legitimate and suspicious or fraudulent billing activities; (iv) assess and interpret how well the models perform, and provide clear explanations for the results; and (v) derive actionable insights or rules from the models outputs that can inform internal controls and policy decisions in the telecommunication industry.

## II. REVIEW OF RELATED WORKS

### A. Fraud in Call Duration

Jabbar and Suharjito (2020) presented fraud detection analysis on call detail record in telecommunication using machine learning. The model adopted in this study was unsupervised learning clustering on call detail records. The study used input variables like call duration, call number, fee, and destination. The clustering identified anomalous call patterns which is related to suspicious call activity. The study reported that the unsupervised method was able to detect fraudulent calls with accurate results when compared with fraud labeled records. The result of the study indicated that there was no strong focus on fraud prevention but only detecting the fraudulent calls after it had occurred in the system.

### B. Fraud in Call Frequency

Hu et al. (2021) conducted a study on mining mobile network fraudsters with augmented graph neural networks. The main objective of the study was to detect fraudulent subscribers in mobile networks by using both user behavior and network structure. The study used a graph neural network to detect fraudulent mobile subscribers by presenting users as nodes using call-based features. Their study found out that traditional Graph Neural Network (GNN) and machine learning approaches can identify fraudulent users accurately. The study only focused on analyzing and detecting the fraud in groups and does not show how this fraud could be prevented in real-time.

Similarly, Zhao et al. (2018) presented a study on detecting telecommunication fraud by understanding the content of a call. They used fuzzy-logic modelling of user behaviour to classify fraud versus legitimate use. Their main objective was to use fuzzy rules built from behavioural features including incoming and outgoing call

ratio, and irregular patterns indicating fraud in the system. Their study found out that fraudulent subscribers produce irregular calling patterns and this can be identified through fuzzy-logic analysis. Fuzzy logic approach only helps with uncertainty; they need richer or enough labelled data and do not directly address real-time prevention of fraudulent activities.

**C. Fraud in Billing Amount or Call Cost**

Osundare et al. (2023) analysed the application of machine learning in detecting fraud in telecommunication-based financial transactions. Their study examined machine learning techniques for detecting anomalies in the billing and underbilled international calls. They used supervised anomaly detection and statistical analysis of billing amount and call detail record features. They found out that under-billed international calls are a strong indicator of bypass and revenue leakage fraud. Their study only focused on detection after the fraud had taken place in the system. It does not integrate real-time blocking or automated prevention measures.

**D. Fraud in Account Age and SIM Activation Date**

Ekwonwune et al. (2022) conducted analysis of Global System for Mobile communication (GSM) subscription fraud detection system. The objective of the study was to determine whether newly registered or recently activated SIM cards display different behavioural pattern that can help in fraud detection. They found out that fraudulent SIM cards are often associated with accounts that were registered recently and with incomplete registration details. The study focused on descriptive analysis and did not implement advanced machine learning models to enhance prediction accuracy.

**E. Machine Learning in Fraud Detection**

Edozie et al. (2025) analyzed artificial intelligence advances in anomaly detection for telecom networks. The main objective of the study was to explore how artificial intelligence can be used to detect anomaly in telecommunication networks in order to replace rule-based methods. They used a systematic review of recent artificial intelligence techniques such as deep learning, hybrid models, and emerging models applied to network traffic and security. The findings emphasized that artificial intelligence outperformed traditional approaches, with deep and hybrid learning models detecting fraudulent activities more accurately while emerging techniques show solid potentials for privacy-sensitive networks. Their study only detects visible fraud anomalies and relies on limited dataset; it is not tested for real-time use, and needs strong computing resources.

Bansal et al. (2024) analyzed fraud detection in the era of artificial intelligence. The main objective of the study was to explore how artificial intelligence and machine learning can strengthen fraud detection in the digital economy. Their study found out that artificial intelligence and machine learning techniques are very reliable in detecting new and complex fraudulent activities. Their study needs strong data infrastructure and technical expertise to implement and might not be suitable for smaller telecommunication companies especially in developing countries.

## III. MATERIALS AND METHODS

**A. Software and Hardware Materials**

The model is implemented using Python programming language which is widely used in data science and machine learning. The following Python libraries are employed in the study: Pandas and NumPy for data manipulation and numerical computation; Scikit-learn for logistic regression modelling and evaluation; Imbalanced-learn for class imbalance using Synthetic Minority Oversampling Technique (SMOTE); and Matplotlib and Seaborn for visualization of data distributions and performance metrics. The hardware used was a personal computer with 16GB RAM, Intel Core i7 processor, and 512GB SSD storage.

**B. Research Design and Data Source**

This study employs a quantitative research design, since it relies on numerical dataset analysis and statistical modelling to develop and evaluate how telecommunication billing system fraud can be prevented (Asenahabi, 2019). The dataset used is a simulated billing dataset obtained from Kaggle, an open-source telecommunication repository. The dataset comprises 1000 Call Detail Records (CDRs). The use of simulated data ensures ethical compliance and removes confidentiality concerns associated with live operator data. The target outcome is binary: 1 = Fraudulent billing transaction and 0 = non-fraudulent billing transaction.

**C. Types of Machine Learning Applied**

Supervised learning is used in this study. In supervised learning, an algorithm is trained using a labelled dataset where input data is paired with its correct outcome. This approach allows the algorithm to learn strong relationships between inputs and outputs, ultimately enabling it to make accurate predictions on new, previously unseen data (Verma et al., 2022).

Logistic regression is a go-to method for binary classification tasks. What it actually does is predict the likelihood of an input belonging to a specific class. The model relies on the logistic function, which turns predicted values into probability estimates, with outputs falling between 0 and 1. Decision trees are a widely used machine learning technique that can handle both classification and regression tasks. To further boost accuracy, decision trees can be combined into a single powerful model called a random forest, which makes its prediction by averaging the individual trees' outputs in the case of regression or by taking the majority vote in classification (Becker et al., 2023). Gradient Boosting builds trees sequentially, where each tree corrects the errors of the previous one, making it a strong ensemble method for tabular data.

**D. Data Preprocessing Steps**

Data preprocessing is used in the study to avoid wrong predictions, improve model performance and reduce difficult computation. The key steps are: Data cleaning (data duplicates are identified and removed); Data transformation (the data is encoded and formatted in a way the machine model can understand); Data integration (the billing records such as call records, SMS records and data usage records are merged); Feature scaling (the variables are adjusted into the same range); and Data splitting (the billing dataset is divided with 80% for training and 20% for testing).

Feature engineering is the process of creating new features from existing billing data records. The engineering features include: Average Call Cost per Day (daily average of billing amount per subscriber); International Call Ratio (proportion of international to total calls); Location Change Frequency (number of cell tower changes per hour/day); Call Duration Variance (variability in call length across calls); SIM/Device ID Change Rate (frequency of SIM swaps or IMEI mismatches); and Peak vs Off-Peak Ratio (proportion of calls made during peak billing hours).

**E. Addressing Class Imbalance**

The dataset was unbalanced towards the majority class because of a common problem in imbalanced datasets. The Random OverSampler method was used to resample the minority class in the training data, which represents both the features and the fraud labels. This method created a new resampled training set. The minority fraud cases increased from 86 to 714 because the majority class had 714 non-fraud cases. This made the dataset more balanced and the data oversampling increased the dataset size from 800 to 1,428 in size.

**F. Model Training and Evaluation**

Three models are trained and evaluated: Logistic Regression, Random Forest, and Gradient Boosting. For each model, the .fit() function was used to teach the models how to find relationships between the features and the fraud labels. After training, the .predict() function was used to make predictions about new or unseen data. The performance metrics used were: Accuracy, Precision, Recall (Sensitivity), F1 Score, and Confusion Matrix.

**Table I: Summary of Review of Related Works**

| S/N | Author, Year | Objectives | Methods | Major Findings | Research Gaps |
|---|---|---|---|---|---|
| 1 | Jabbar & Suharjito, 2020 | Evaluate unsupervised learning for detecting suspicious call activities | K-means and DBSCAN clustering on CDRs | Unsupervised method detected fraudulent calls with accurate results | No focus on fraud prevention; only detection after fraud occurred |
| 2 | Hu et al. 2021 | Detect fraudulent subscribers using user behavior and network structure | Graph neural network on call-based features | GNN and ML can identify fraudulent users accurately | Only group-based detection; no real-time prevention |
| 3 | Zhao et al. 2018 | Classify fraud vs. legitimate use using fuzzy rules | Fuzzy sets from behavioral features | Fraudulent subscribers produce irregular calling patterns | Needs richer labelled data; no real-time prevention |
| 4 | Osundare et al. 2023 | Detect anomalies in billing and underbilled international calls | Supervised anomaly detection and statistical analysis | Under-billed calls are strong indicators of fraud | Only detection; no real-time blocking or prevention |
| 5 | Djomadji et al. 2023 | Identify SIM box bypass fraud using CDR features | Random Forest, SVM and XGBoost | Fraudulent activity dominated by high outgoing calls | Only detection; no real-time interpretable prevention |
| 6 | Edozie et al. 2025 | Use AI to detect anomaly in telecom networks | Systematic review of deep learning and hybrid models | AI outperformed traditional approaches | Relies on limited dataset; not tested for real-time use |

*Source: The researcher (2025).*

## IV. RESULTS AND DISCUSSION

### A. Data Presentation and Analysis

An analysis was conducted using a simulated dataset which contained billing records of customers. The dataset comprised of 1000 billing records which represented transactions of different customers. The variables used reflected the key fraud indicators in the literature review. Table II shows the dataset features, descriptions, and data types.

**Table II: Dataset Features and Parameters**

| Parameters | Description | Data Type | Example Value |
|---|---|---|---|
| Call Duration | Length of call in minutes/seconds | Quantitative | 180 Seconds |
| Call Frequency | Number of calls made per day/week | Quantitative | 15 calls/day |
| Timestamp | Time and date of call/billing transaction | Date/time | 2025-07-10 14:23 |
| Call Destination | National or international | Qualitative | International |
| Billing Amount | Cost of charge/call | Quantitative | ₦100 |
| Account Age | Days since SIM was registered | Quantitative | 45 days |
| Device ID Consistency | Indicator of SIM/IMEI match | Binary | 1 |
| Data Usage Pattern | Amount of average data used daily | Quantitative | 2.5GB/day |
| Location Change | Frequency of switching locations | Quantitative | 4 changes/day |
| Incoming/Outgoing Call Ratio | Ratio of incoming to outgoing calls | Quantitative | 0.67 |
| Fraud Label | Fraudulent (1) or Non-fraud (0) | Binary | 0 |

*Source: The researcher (2025).*

### B. Model Performance Before Oversampling

Logistic regression, random forest, and gradient boosting models were evaluated. All models produced high accuracy values but low recall values for the fraud class because of the imbalance. The model performance before oversampling is presented in Table III.

**Table III: Model Performance Before Oversampling**

| Model | Accuracy | Recall (Fraud) | F1 Score (Fraud) |
|---|---|---|---|
| Logistic Regression | 0.8900 | 0.00 | 0.00 |
| Random Forest | 0.8850 | 0.00 | 0.00 |
| Gradient Boosting | 0.8800 | 0.00 | 0.00 |

*Source: The researcher (2025).*

### C. Individual Model Performance After Oversampling

After random oversampling was applied to the training data, all the models were retrained. The individual model performances are presented in Tables IV, V and VI.

**Table IV: Logistic Regression Performance Before and After Oversampling**

| Metrics | Before Oversampling | After Oversampling |
|---|---|---|
| Accuracy | 0.8900 | 0.5500 |
| Precision (1) | 0.00 | 0.11 |
| Recall (1) | 0.00 | 0.45 |
| F1-Score (1) | 0.00 | 0.18 |

*Source: The researcher (2025).*

**Table V: Random Forest Performance Before and After Oversampling**

| Metrics | Before Oversampling | After Oversampling |
|---|---|---|
| Accuracy | 0.8850 | 0.8500 |
| Precision (1) | 0.00 | 0.00 |
| Recall (1) | 0.00 | 0.00 |
| F1-Score (1) | 0.00 | 0.00 |

*Source: The researcher (2025).*

**Table VI: Gradient Boosting Performance Before and After Oversampling**

| Metrics | Before Oversampling | After Oversampling |
|---|---|---|
| Accuracy | 0.8800 | 0.7550 |
| Precision (1) | 0.00 | 0.09 |
| Recall (1) | 0.00 | 0.14 |
| F1-Score (1) | 0.00 | 0.11 |

*Source: The researcher (2025).*

**D. Comparative Model Performance After Oversampling**

After oversampling was carried out, the sensitivity of the models to detect fraudulent patterns improved. Logistic regression showed the most balanced improvement. The comparative model performance after oversampling is presented in Table VII.

**Table VII: Comparative Model Performance After Oversampling**

| Model | Accuracy | Precision | Recall |
|---|---|---|---|
| Logistic Regression | 0.5500 | 0.11 | 0.45 |
| Random Forest | 0.8500 | 0.00 | 0.00 |
| Gradient Boosting | 0.7550 | 0.09 | 0.14 |

*Source: The researcher (2025).*

**E. ROC Curve Analysis**

Logistic regression showed an improved distinction between fraud and non-fraud cases after oversampling. Before oversampling, the model achieved an AUC of 0.46, which means the model performed below random guessing because of high imbalance between fraudulent and non-fraudulent records. After random oversampling was applied, the AUC value increased from 0.46 to 0.73 which showed a very good improvement in the model's ability to be able to tell the difference between both classes. This means that oversampling enhances the model's sensitivity and overall accuracy to detecting fraudulent transactions which makes the logistic regression model suitable for fraud prevention in real world application.

**F. Confusion Matrix Analysis**

After training the three models and also assessing their first performance results, the confusion matrix was used. To better understand, the confusion matrix was visualized using the heatmap diagram. The matrix quantifies the number of: True Positives (TP): fraudulent transactions correctly predicted as fraud (0 cases); True Negatives (TN): non-fraudulent transactions correctly predicted as genuine (178 cases); False Positives (FP): genuine transactions incorrectly predicted as fraud (0 cases); and False Negatives (FN): fraudulent transactions incorrectly predicted as genuine (22 cases). The analysis from the feature relationships showed that long call duration and high data consumption contributes largely to the possibilities of billing anomalies.

**G. Discussion**

The analysis reveals several key findings on the detection of billing anomalies for revenue assurance and fraud prevention. The initial evaluation of the three models (Logistic Regression, Random Forest, and Gradient Boosting) showed overall accuracy. However, after close analysis of the classification reports and confusion matrix, the models performed poorly and found it difficult to identify fraudulent cases of the minority class which was evident in low precision, recall, and F1-scores and also a high number of false negatives. This was as a result of imbalanced dataset.

To address this concern, random over sampling was carried out to train the data. The later retraining and analysis of the three models produced different results from the previous results. While Logistic Regression showed improvement in recall for the minority class which demonstrated a better ability to detect potential frauds, it showed low precision which indicated the probability of the model flagging non-fraudulent transactions as fraudulent transactions. Random Forest and Gradient Boosting maintained a high accuracy but the models could not still effectively identify the minority class.

These results showed that while oversampling can be a good technique to improve on the model's ability to detect more fraudulent transactions (minority class recall), it may not always result in a balanced improvement across the relevant metrics especially the precision, when dealing with imbalanced datasets. The class imbalance of initial results aligns with the work of Terzi et al. (2021) and Pratihar et al. (2023) which was on telecommunication fraud detection. Their findings confirm that when the data is imbalanced, the model will tend

to favour the majority fraud class and fail to detect the real fraud and because of this real fraud cases are missed. The trade-off between precision and recall is a key factor in anomaly detection, where minimizing false negative (missing the actual fraud cases) is most important. This aligns with the study of Verma et al. (2022), who confirms that sampling techniques are very effective in order to reduce false negatives.

## V. SUMMARY, CONCLUSION AND RECOMMENDATIONS

### A. Summary

This study highlights how machine learning models can be developed and used to prevent fraudulent transactions in the telecommunication billing systems. A simulated dataset that consisted of 1000 billing records of customers' transactions was used. Three machine learning models were trained: Logistic Regression, Random Forest, and Gradient Boosting. Their performances were compared before and after oversampling using accuracy metrics. The results showed that the Logistic Regression model performed better after oversampling. The study analysis showed that a properly trained machine learning model can help telecommunication operators prevent fraudulent transactions in billing systems, and also reduce revenue loss.

### B. Conclusion

The findings in this study showed that machine learning model development can have a lasting positive impact on the billing system fraud prevention in telecommunication billing systems with a well-structured framework in place. The Logistic Regression model is seen to be a practical option for preventing billing fraud because it is easy to interpret and understand and this is important for decision-making in real-time systems. The study concludes that telecommunication operators can prevent fraudulent transactions in their billing systems when they study past billing records, learn from the records and quickly detect unusual anomalies in order to help them make informed decisions by either demanding for authentication or blocking the fraudulent transactions immediately before it causes revenue loss to the system.

### C. Contributions to Knowledge

This study contributes to the already existing body of knowledge on telecommunication billing fraud detection by developing a framework for using machine learning techniques to prevent fraud from happening rather than just detecting the billing fraud without a proactive measure in place. This study also contributes to knowledge by providing practical evidence that data balancing improves model performance when dealing with fraud related datasets. This study contributes to already existing knowledge by identifying logistic regression as a suitable model technique for preventing fraudulent transactions in real-time telecommunication billing system.

### D. Recommendations

The problem with working with an unbalanced dataset can reduce the model's ability to detect billing anomalies. As a result, further research should explore other resampling techniques or combining techniques, such as undersampling the majority class or using more advanced methods. Future study should also acquire more data especially dataset with minority class which could further improve the model training and performance. While logistic regression technique is seen to be accurate, other techniques like Support Vector Machines which are known for handling imbalanced datasets better should be used. This study used a simulated dataset for the analysis; future study should use real billing datasets from telecommunication operators to test model performance in real-world conditions. Future studies can also explore deep learning approaches such as neural networks to help detect and prevent complex fraud behaviors.

## REFERENCES

[1] Asenahabi, B. M. (2019). Basics of research design: A guide to selecting appropriate research design. International Journal of Contemporary Applied Researches, 6(5): 2308–1365.

[2] Aziz, Z. and Bestak, R. (2024). Insight into anomaly detection and prediction and mobile network security enhancement leveraging K-means clustering on call detail records. Sensors, 24(6): 1710–1716.

[3] Balmer, R. E., Levin, S. L. and Schmidt, S. (2020). Artificial intelligence applications in telecommunications and other network industries. Telecommunications Policy, 44(6): 101970–101977.

[4] Bansal, U., Bharatwal, S., Bagiyam, D. S. and Kismawadi, E. R. (2024). Fraud detection in the era of AI: Harnessing technology for a safer digital economy. Advances in Finance, Accounting, and Economics, 7: 143–164.

[5] Becker, J. et al. (2023). Random forest for classification and fraud detection in financial data. Journal of Machine Learning Research, 24(1): 1–30.

[6] Bello, I. A. et al. (2023). Real-time telecommunication fraud detection using machine learning and stream processing. Telecommunications Policy, 47(3): 102456–102465.

[7] Djomadji, E. M. T. et al. (2023). Machine learning-based approach for identification of SIM box bypass fraud in a telecom network based on CDR analysis. Journal of Big Data, 10(1): 1–21.

[8] Edozie, C. et al. (2025). Artificial intelligence advances in anomaly detection for telecom networks. IEEE Access, 13: 12001–12020.

[9] Ekwonwune, E. N. et al. (2022). Analysis of global system for mobile communication (GSM) subscription fraud detection system. International Journal of Computer Science, 14(2): 45–58.

[10] Hu, X. et al. (2021). Mining mobile network fraudsters with augmented graph neural networks. IEEE Transactions on Network and Service Management, 18(3): 3832–3845.
[11] Hu, X. et al. (2023). Deep learning for telecommunications fraud detection: A survey. IEEE Communications Surveys and Tutorials, 25(2): 1100–1125.
[12] Jabbar, M. A. and Suharjito (2020). Fraud detection analysis on call detail record in telecommunications using machine learning. International Journal of Advanced Science and Technology, 29(5): 3764–3778.
[13] Johnson, R. et al. (2022). Explainability in machine learning for fraud detection: Challenges and opportunities. Expert Systems with Applications, 198: 116–124.
[14] Love, P. E. D. et al. (2023). Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. Advanced Engineering Informatics, 57: 102017–102024.
[15] Nguyen, T. T. et al. (2020). Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey. Artificial Intelligence Review, 52(1): 77–124.
[16] Osundare, O. S. et al. (2023). Application of machine learning in detecting fraud in telecommunication-based financial transactions. African Journal of Computing and ICT, 16(2): 33–45.
[17] Psychoula, I. et al. (2021). Explainability in machine learning for banking and financial services. IEEE Access, 9: 89506–89521.
[18] Terzi, D. S., Sagiroğlu, S. and Kilinç, H. (2021). Telecom fraud detection with big data analytics. International Journal of Data Science, 6(3): 191–204.
[19] Verma, A. et al. (2022). Supervised machine learning approaches for fraud detection: A review. Procedia Computer Science, 215: 671–680.
[20] Xu, Y. et al. (2023). Real-time fraud detection in billing systems: A streaming approach. Future Generation Computer Systems, 140: 300–310.
[21] Yehya, A. and Salhab, H. (2023). Telecommunications fraud detection using machine learning, mainly Support Vector Machine (SVM). International Journal of Advanced Computer Science and Applications, 14(4): 211–220.
[22] Zhao, J. et al. (2018). Detecting telecommunication fraud by understanding the contents of a call. Cybersecurity, 1(1): 1–10.