Journal of Global Research in Electronics and Communication
Volume 1, No. 11, November 2025
Available Online at: www.jgrec.info
RESEARCH PAPER

An Intelligent System for Identifying Fraud Phone Calls Using Machine Learning Algorithms

Sandeep Gupta
SATI, Vidisha
Sandeepguptabashu@gmail.com

***Abstract*—Phone fraud, also known as spam and unwanted calls, is a major problem in the telecom industry, costing millions of dollars annually all around the globe even as technology goes further and further ahead. The rise of phone scams, also known as spam using a mobile phone or a telephone, has become a common security risk for businesses and individuals alike. Machine learning and artificial intelligence have shown promising results in analyzing and detecting fraudulent or harmful phone calls. This paper gives an efficient method to forecast fraud calls through Call Detail Records (CDR) based on a selection of different machine learning and clustering algorithms. Data cleaning, normalization, feature selection, and data balancing based on SMOTE were deemed applicable to the CDR dataset that includes fundamental features, including caller ID, called ID, duration, cost, destination, and call type. The various models that have been used include DB scan, SVM, GCN, and XGBoost which revealed patterns of fraudulent behavior. The proposed XGBoost model had the best accuracy score of 96.7 %, illustrating that it could better recognize fraud than the others.**

Keywords*—Telecommunications, Artificial Intelligence, Fraud Call Detection, Machine Learning, Telecommunication Security.

I. INTRODUCTION

Fraud and spam calls via phone are a constant and continuously developing threat to people, organizations, and government. An estimated three billion dollars was lost in 2021 as a result of the millions of fraud accusations filed in the US (more than three million according to the FTC alone) [1][2][3]. Spoofing, impersonation, and digital manipulation are some of the advanced methods that fraudsters use to victimize people in order to obtain sensitive data, steal funds, or discredit images. Such fraudulent activities have an effect not only on the financial side but also on psychological effects where people are subjected to stress and anxiety [4].

The extent of telecom fraud on a global basis is shockingly alarming as the sum of the unsolicited calls made every day in different regions totals to millions. These attacks are dependent upon holes that can exist in both caller ID systems and protocols of the network and as such blacklisting or any type of hard coded mechanisms of rules unable to sniff out these types of attacks [5][6][7]. Naturally, the modus operandi of these fraudulent calls is to induce urgency, fear, or emotive appeal in their victims that may make them act in haste at great costs to them and their organizations security wise.

Manually fetching and examining call records and audio samples of the call used to be the main factor that would allow detecting fraudulent calls traditionally [8]. Such rule-based networks remain costly, time-consuming, and inefficient, as unseen and versatile fraud patterns are found. In addition, manual solutions cannot scale and do not have the flexibility to mitigate new scammers in real-time [9][10]. The telecommunications industry which is currently experiencing a rising number of fraud cases, has been losing substantial revenue and harming its relationship with its customers through the constraints of the traditional fraud prevention systems. With such challenges, there has been an increased interest in smart and automated systems that aid in the accurate detection of fraudulent phone calls [11][12][13]. An examination of the malicious calls and their metadata, including calls that last, how frequently they occur, the location of the caller, and the transcripts is capable of producing major behavioral patterns and other deceptive tactics of the scammers. Sophisticated technologies like voice recognition and speech-to-text translation may help distinguish the particular content and psychological manipulation strategies applied in the course of such calls [14][15].

Using machine learning algorithms [16][17], it is possible to make fraud detection more efficient and more effective to minimize financial losses and security risks. Machine learning in the fraud call detection has a number of advantages. It is possible to design machine learning algorithms that can learn anomalies and patterns in data that could enable them to detect fraudulent calls better because they can be used to detect them more effectively [18][19]. Also, machine learning models have the ability to learn as they experience the quality of their performance, which generally increases with the accumulation of the model training data [20][21].

A. Significance and Contribution

Phone-based fraud and spam calls is a prominent and developing source of cyber insecurity and has critical impacts on people, institutions and governance at all levels across the globe. Manual inspection and rule-based detection are considered to be effective no longer as the radical attack tricks, like spoofing, identification, and emotional tricks has already become a problem. These methods take advantage of vulnerabilities in the caller identification and communication protocols such that both financial and psychological damages can be large. An automated, intelligent system that might identify and prevent such fraudulent operations in real time is crucial, given the worldwide problem with such activities and the ineffectiveness of present protection systems. The research tackles the most urgent issue of fraudulent phone call detecting through analyzing Call Detail Records (CDR) using machine learning approach to provide a data-driven and scalable solution to the problem in order to contribute to bolstering the telecommunication security and eliminating the losses. This study presents some major contributions to the domain of Identifying Fraud Phone Calls:

- Caller ID, Called ID, Duration, Fee, Destination, and Type data are all part of 11,418 Call Detail Records (CDRs) from customer invoices. This data is used to test calling habits and potential frauds.
- An end-to-end data preprocessing pipeline is applied, such as removing outliers, normalization with min-max normalization, and addressing the problem of the imbalance of classes using SMOTE.
- Feature selection is a technique for improving model performance by focusing on the most useful features.
- XGBoost fraud detection machine learning framework is proposed on identifying the fraudulent phone calls based on the CDRs.
- Combined a number of metrics and statistics, including recall, accuracy, precision, ROC curve, and F1-score, to determine the models' efficacy.

B. Justification and Novelty

The novelty and justification of the research are based on its end-to-end approach of identifying fraudulent phone calls via a Call Detail Record (CDR) database and XGBoost algorithm, which is well-suited to handle structured data and detect non-linear patterns. This paper can be distinguished by handling these important issues like data imbalance, irrelevant features, and outliers' effects by using a preprocessing mechanism that consists of feature selection, min-max normalization, and SMOTE designed to induce synthetic oversampling. In contrast to most of the other solutions available, the current work focuses not only on the data quality but also on how to explain the significance of particular constructs, while it is also powered by the traditional CDR characteristics along with a high-precision classifier. The combination of the ROC-AUC and precision-recall analysis also helps to bring to light the nature of model behavior, and thus, the benefits of the approach extend not just technically but practically to the utilization of telecom fraud monitor systems.

C. Structure of The Paper

The outline of the paper is as follows: Research in Section II is reviewed. This section describes the approach and ML models that were utilized. The results and analysis are presented in Section IV. Section V wraps up the study and provides a framework for what comes next.

II. Literature Review

There are a lot of research works devoted to the problem of fraudulent phone calls identification that have been examined and discussed in order to guide the evolution of the current work and give it a stronger foundation.

Singh, Singh and Singh (2025) provide a unique approach to real-time fraud detection that utilizes RAG technology to tackle this problem from two angles. To begin, there is a policy-checking tool built into the system that is continuously being updated. To ensure an honest and open conversation, it uses RAG-based models to check if the caller isn't trying to steal sensitive information. A real-time user impersonation check that employs a two-step verification method to prove the caller's identity would also be helpful in ensuring responsibility. One key improvement of the system that makes it more flexible is the ability to change policies without having to retrain the whole model. proved RAG-based technique using simulated call recordings, outperforming state-of-the-art methods with a 97.98% accuracy and an F1-score of 97.44% with 100 calls. This robust and flexible fraud detection system is ideal for practical application [22].

Bhargavi and Shivani (2024) discovered and developed 29 features that may be utilized by algorithms trained on machine learning data to forecast potentially harmful phone calls. People and businesses alike are more vulnerable to spam, fraud, and other forms of unsolicited phone calls. A new development in the fight against damaging or fraudulent calls is the growing effectiveness of methods based on machine learning and artificial intelligence. This paper provides a high-level summary of AI-based spam and fraud detection and analysis techniques, as well as a discussion of the problems and possible solutions to those problems. A novel method for detecting fraudulent phone calls is proposed, with the expectation that it can attain very high precision and accuracy rates. According to the findings, the most efficient method was able to cut the number of malicious calls that were previously unblocked by as much as 90 %, while still maintaining a precision rate that was higher than 93.79% for benign call traffic. Furthermore, the results showed that these models could be executed effectively with little latency overhead [20].

Zhao et al. (2024) demonstrate that Chameleon separated signals can be recognized with an average accuracy of 92.69%, surpassing the commonly used Fast ICA method, which only reaches 25% accuracy. To take into consideration the set of changes in the environment, apply the Fréchet Inception Distance (FID) model as a guiding indicator concerning the migration of the model. Moreover, present the Inductive Vector that allows key-identifying model to adjust to the changed environmental conditions like environment, phone location, and the variety of users. The Inductive Vector adjusts the model parameters based on the shift in FID. In scenarios with various phone locations, the Inductive Vector significantly improves recognition accuracy from 61% to 98%, outperforming the best existing keystroke recognition algorithm [23].

Hong, Connie and Goh (2023) offer a tool that utilizes machine learning, more especially deep learning and natural language processing to identify and categorize scam calls. Training the model to detect fraudulent conversations involves feeding it data from a dataset that contains both scam and non-scam calls. In order to create a reliable classifier, they employ various natural language processing (NLP) techniques, such as word embeddings, text preprocessing, and the Google API to convert audio samples to text. In terms of identifying fraudulent calls, the Long Short-Term Memory (LSTM) algorithm outperformed the competition with an accuracy rate of 85.61% [24].

Zhang et al. (2022) accessing the raw data stored on users' mobile devices could potentially breach the ever-tightening private data privacy rules and laws, such as GDPR 71, and can identify mobile phone callers without worrying about their privacy if they use the right statistical approaches to remove private information while keeping personal features. In this study, they employ privacy-preserving mobile data to train a model that can detect and block four types of callers: regular people (in other occupations), cab drivers, food delivery people, telemarketers, and fraudsters. Results from a three-month validation run on an anonymized dataset including 1,282 users in Shanghai City show that the suggested model is capable of achieving an accuracy of 75% or higher [25].

Kale et al. (2021) A blacklist of known fraudulent numbers is essential for detecting phishing calls. This causes issues whenever the system is presented with fresh or unfamiliar numbers. The best solution to this problem would be to implement a system that analyzes the caller-victim conversation in order to detect phishing attempts. Using several machine learning techniques, they analyzed the intent of call transcripts. They built two models and compared them. Both the CNN-based model and the Naive Bayes Algorithm-based model obtained respectable levels of accuracy: 94.57% and 97.21%, respectively [26].

Gowri et al. (2021) the prior data set of spam calls is gathered first. For the purpose of predicting malicious calling, the dataset contains multiple label-based features. To identify the fraudulent calls, their main emphasis is on the Recurrent Neural Network (RNN) method. By analyzing various state-of-the-art machine-learning approaches with the proposed characteristics, they deduce that the best solution may reduce fraudulent calls to 90% while maintaining an accuracy of over 90% for binary calls [27].

Machine learning, deep learning, and privacy-preserving approaches have made great advances in fraud call detection in previous research, but there are still many important problems that have not been solved. Many models rely heavily on static features, predefined blacklists, or limited contextual understanding, which reduces their adaptability to evolving scam techniques. Furthermore, approaches like Naive Bayes, LSTM, and CNN achieve high accuracy but often lack real-time capabilities, policy adaptability, or dynamic fraud pattern recognition. Additionally, most models do not incorporate techniques for handling imbalanced datasets effectively or fail to address the high false positive rates that can undermine user trust. To bridge these gaps, this study proposes a robust and adaptive fraud detection framework using XGBoost on Call Detail Records (CDR), incorporating comprehensive preprocessing including SMOTE for data balancing, min–max normalization, and focused feature selection. This solution emphasizes scalable deployment, reduced manual intervention, and real-time fraud detection capabilities by learning behavioral patterns from CDR data, offering a high-performance, cost-effective, and practical approach for modern telecom fraud detection systems.

Table I presents a summary of recent studies on Fraud Phone Calls Prediction, highlighting innovative models, datasets used, major findings, and the challenges faced.

TABLE I. COMPARATIVE ANALYSIS OF RECENT STUDIES ON PREDICTIVE MODELING OF FRAUD PHONE CALLS PREDICTION USING MACHINE LEARNING

Author	Proposed Work	Dataset	Key Findings	Challenges/recommendations
Singh, Singh and Singh (2025)	Real-time fraud detection using RAG with policy checks and impersonation detection	Synthetic call recordings (100 calls)	Achieved 97.98% accuracy and 97.44% F1-score; highly adaptable and effective in real-time environments	Depends on real-time transcription quality and dynamic policy update integration
Bhargavi and Shivani (2024)	ML-based fraud detection using 29 identified features	Dataset with 29 engineered features	Reduced malicious calls by 90%, precision over 93.79%, and low latency	Feature design is crucial; it needs efficient deployment in real-world conditions
Zhao et al. (2024)	Chameleon signal separation with Inductive Vector guided by FID	Signal data with environment variables	Accuracy improved from 61% to 98%; outperformed Fast ICA (25% accuracy)	Needs adaptation for environmental changes; relies on FID for tuning
Hong et al. (2023)	NLP + Deep Learning (LSTM) based scam call detection	Scam/non-scam call dataset with Google API transcription	The LSTM model achieved 85.61% accuracy	Dependent on speech-to-text quality and NLP context handling
Zhang et al. (2022)	Caller identity prediction using privacy-preserving mobile data	Anonymized data from 1,282 users over 3 months (Shanghai)	Achieved over 75% accuracy without violating privacy	Balancing privacy protection with model performance
Kale et al. (2021)	Fraud call classification via transcript intent analysis	Conversation transcripts	CNN achieved 97.21% accuracy; Naive Bayes reached 94.57%	Effective for unseen numbers; needs robust context modeling
Gowri et al. (2021)	Spam detection with RNN using labeled call features	Historical spam call dataset	Reduced malicious calls by 90%, accuracy maintained above 90%	Efficient deployment is possible with minimal latency; it depends on label quality

III. RESEARCH METHODOLOGY

The methodology of this study centers around analyzing Call Detail Records (CDR) collected from customer billing data. Data cleaning, missing value handling, and outlier removal were all part of the preprocessing stages to guarantee quality. Using the min-max scaling, the data was normalized. Synthetic minority samples were generated using the feature selection, SMOTE approach. A split into a training set and a testing set was the next step in preparing the dataset. Finally, the proposed XGBoost model was implemented, which enhanced the understanding of call behaviour for future research or fraud detection. Using standard metrics like accuracy, precision, recall, F1-score, and ROC curves, the model was found to accurately predict and categorize fraudulent phone calls. Figure 1 shows the entire process.

The following is a comprehensive breakdown of the proposed flowchart for identifying fraudulent phone calls.

A. Data Collection

Customer billing records should be searched for the Call Detail Records (CDR) data for the usage period beginning May 1, 2018, and ending May 31, 2018. A large number of attributes were present in the 11,418 rows that made up the acquired data. The researchers settled on six characteristics to serve as independent variables: number (Caller ID), number (Called ID), length, cost, destination, and kind. Data visualizations such as bar plots and histograms were used to examine feature correlations, etc., and are given below:

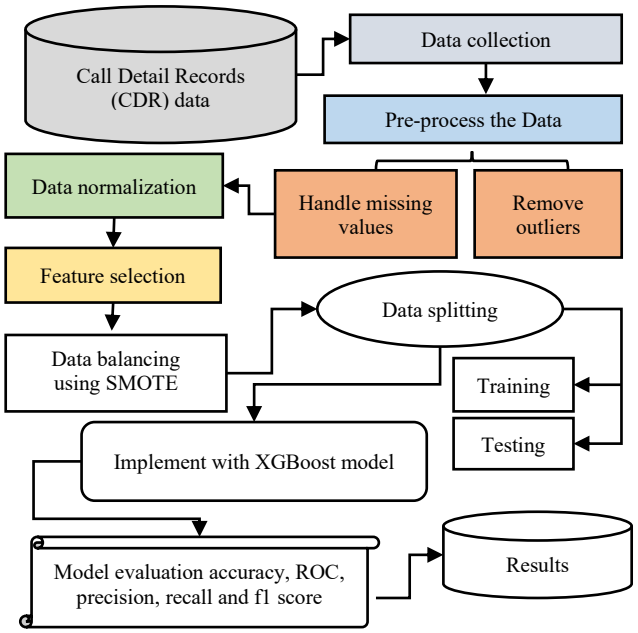

Fig. 1. Proposed flowchart for Fraud Phone Calls Prediction

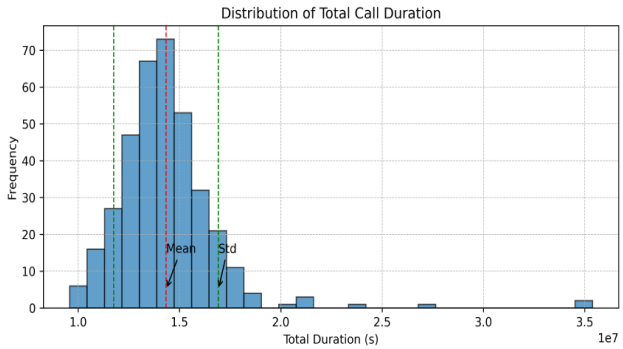

Fig. 2. Histogram for Data distribution

Figure 2: The histogram displays the distribution of total call duration in seconds, showing a right-skewed pattern. Most of the data is concentrated between approximately 1.0×10^7 and 1.8×10^7 seconds, with a peak (mode) around 1.45×10^7 seconds. The mean is marked with a solid blue line, while the standard deviation boundaries are indicated by dashed green lines. A few extreme values on the right suggest the presence of outliers. The visualization effectively highlights the central tendency and variability of the call durations, indicating that most call durations cluster near the mean, with a gradual decline toward higher values.

B. *Data Pre-Processing*

The data preparation process began with collecting Call Detail Records (CDR), followed by concatenating and cleaning the dataset, and extracting relevant features. Data processing was carried out through the elimination of outliers and missing values. After this, transformation and normalization of data was carried out. Preprocessing The major preprocessing steps are the following:

- **Remove Missing Value:** Removing missing values, also known as handling missing data, involves strategies to address incomplete information within a dataset. The procedure to use is determined by the size and type of the lost data and the purpose of the analysis.
- **Remove Outliers:** Removal of outliers is just the elimination of data that lies far outside of a given body of data. It is essential to data analysis and machine learning because outliers may distort the methods of statistics, and they detrimentally affect the performance of models.

C. *Data Normalization*

Normalization of records was performed on the basis of the min- max technique to restrict the values within the range of 0 and 1. This was done with an objective of maximizing the functioning of the classifiers employed and the influence of outliers. Normalization was done as highlighted by the following mathematical Equation (1):

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

where X is the original feature value, X' is the normalized value, X_{min} is the lowest value of the feature and X_{max} is the highest value of the same.

D. *Feature Selection*

Feature selection refers to the method that involves the selection of a subset of the attributes of a larger set of available attributes to deliver better performances of machine learning models at the lowest computational costs and with better interpretability. It can be done by discovering and removing irrelevant or redundant attributes and concentrating the model to be concerned with the most informative aspects of the information. Machine learning feature selection involves the process of selecting a subset (and possibly a small subset) of relevant features of an original set of features with the purpose of constructing a model.

E. *Data Balancing using Synthetic Minority Over-sampling Technique (SMOTE)*

Data balancing is a technique used in ML, particularly in classification tasks, to address class imbalance, where one or more classes have significantly fewer samples than others. SMOTE is a widely used oversampling method for addressing class imbalance in datasets.

F. *Data Splitting*

This model's efficacy may be evaluated by dividing the dataset into two parts: training and testing. It used 80% of the data to train the model, and 20% to test and evaluate its performance.

G. *Classification with XGBoost Model*

The acronym for "eXtreme Gradient Boosting" is XGBoost. Because of its fast execution and good model performance, XGBoost is mainly used for this purpose. XGBoost aggregates the results of multiple algorithms into a single model through the use of ensemble learning techniques. When it comes to memory efficiency, XGBoost has covered, and it also works with distributed and parallel computing [28]. To further improve classification accuracy and save running time, XGBoost can automatically utilize the multi-threaded CPU for calculations. Here is a quick rundown of the algorithm:

The model prediction function for k decision trees is given by Equation (2):

$$\hat{y} = \sum_{K=1}^{K} f_k(x_i) F = \{f(x) = w_{q(x)}\} q: R^m \to T, w \in R^T \quad (2)$$

Here, $w_{q(x)}$ is the proportion of the leaf node q, f (x) is the regression tree, xi is the i-th input sample, and F is the hypothesis space [29]. The target function is expanded using Taylor's second-order formula once the t-tree is generated, and then the constant term is eliminated to obtain. It can be found in Equation (3):

$$\tilde{L}^{(t)} = \sum_{i=1}^{n}[g_i f_i\left(x_i + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \quad (3)$$

Utilizing the initial component of the aforementioned algorithm, one may approximate the disparity between the anticipated and actual scores.

H. Evaluation Metrics

The proposed design was evaluated using a wide variety of performance criteria. By comparing the expected outcomes predicted by the trained models with the observed values, the True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) were determined. A number of important evaluation metrics were computed using these data; these include F1-score, recall, accuracy, and precision.

1) Accuracy

The fraction of the total occurrences in the dataset (input samples) that were correctly predicted by the trained model. Equation (4) is expressed as follows:

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \quad (4)$$

2) Precision

Precision is defined as the percentage of positive occurrences that the model accurately predicts relative to the total number of positive occurrences. Precision indicates. How good the classifier is in predicting the positive classes is expressed as Equation (5).

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \quad (5)$$

3) Recall

This metric, the ratio of events that were accurately predicted as positive to all instances that should have proved positive. In mathematical form it is given as Equation (6).

$$Recall = \frac{\text{TP}}{TP+FN} \quad (6)$$

4) F1 score

It aids in maintaining a healthy equilibrium between recall and precision by combining the two concepts of the harmonic mean. Its range is [0, 1]. Mathematically, it is given as Equation (7).

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Receiver Operating Characteristic Curve (ROC): The receiver operating characteristic (ROC) plots the ratio of the number of cases that are correctly and incorrectly classified as positive, given a set of decision cut-off points. TPR is commonly referred to as sensitivity or recall, while FPR is equal to 1-specificity. This is where TP stands for True actual and true predicted, FP for False actual and true forecasted, TN for False actual and false predicted, and FN for True actual and false predicted.

IV. Results and Discussion

This section describes the experimental setup and how well the proposed model was trained and tested. The paper's server is a Windows machine with Python 3.8 installed, a GeForceRTX3090GPU, a 16-core Intel(R) Xeon(R) Gold 5218 CPU running at 2.30GHz, and all models are executed on this server. Important performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the proposed model, and the results are displayed in Table II. Call Detail Records (CDR) were used to train the model. How well the XGBoost model predicted fraudulent phone calls using data from Call Detail Records (CDRs). With an impressive 96.7% accuracy rate, the model clearly performed admirably when it came to calling types. The fact that it achieved a recall of 81.8% and a precision of 47.3% indicates that a considerable number of the calls that were marked as fraudulent were actually false positives. Overall, performance in terms of the precision-recall trade-off is moderate, as indicated by the F1-score of 60.

TABLE II. Experiment Results of Proposed Models for Fraud Phone Calls Prediction On CDR Data

Performance Matrix	XGBoost
Accuracy	96.7
Precision	47.3
Recall	81.8
F1-score	60

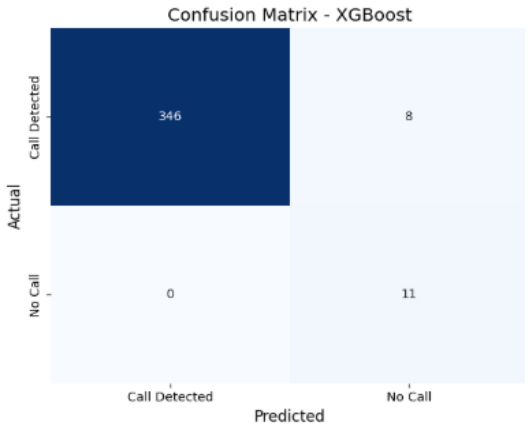

Fig. 3. Confusion matrix for XGB model

Figure 3 shows the model's performance in a classification task with two classes: "Call Detected" and "No Call." The confusion matrix represents this problem. This data set displays the expected classes in columns and the actual classes in rows. As seen in the matrix, 346 occurrences were correctly classified as "Call Detected" and 11 occurrences were accurately classified as "No Call." The model accurately predicted 8 cases of "Call Detected" when the actual class was "No Call," while it failed to identify 0 cases of "Call" when one was present; these are known as false negatives.

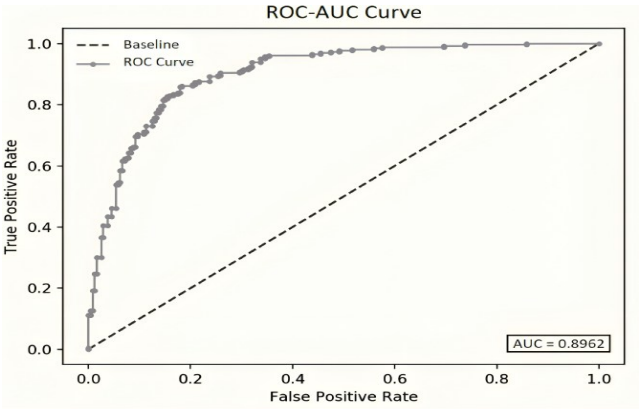

Fig. 4. Precision-Recall analysis of the XGB model

The XGBoost model's performance is shown by the ROC-AUC curve in Figure 4. On one side of the graph is the False Positive Rate, while on the other is the True Positive Rate (sensitivity). The ROC curve indicates great classification capabilities with a strong increasing trend towards the top-left corner. At 0.8962, the AUC shows that the model is very accurate. Confirming the model's efficacy in recognizing calls with few false positives, an AUC close to 1.0 indicates great discriminating between classes.

A. Comparative Analysis

A comparison accuracy plot was made with other available models, as represented in Table III. Using data from Call Detail Records (CDRs), this study compares the accuracy of different prediction algorithms used to identify fraudulent phone calls. With a result of 91.08% and 91.08%, respectively, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) model is marginally inferior to the Support Vector Machine (SVM). The Graph Convolutional Network (GCN) model also improved due to its good model accuracy of 93.4%. It is noteworthy to mention that the proposed XGBoost model produced the highest approximate value of 96.7% which shows that it is highly competent in detecting and identifying the activity of fraudulent calls.

TABLE III. ACCURACY COMPARISON OF DIFFERENT PREDICTIVE MODELS OF FRAUD PHONE CALLS PREDICTION USING CDR DATA

Performance Models	Accuracy
DBSCAN [30]	90.39
SVM [31]	91.08
GCN [32]	93.4
XGBoost	96.7

A significant advantage in identifying fraud phone calls is the high accuracy of the suggested XGB model, which accurately distinguishes between fraudulent and legitimate calls. This good accuracy guarantees reliable performance in real-time situations, whereby it is important to reduce false classification. Moreover, being unsupervised, there is no requirement for labelled data; hence, it is very efficient and can work on a massive amount of data, such as Call Detail Records (CDR). The fact that the model can understand patterns and cluster similar behaviors increases its viability to detect anomalies, which offers a reliable and easy-to-implement solution in detecting fraud in computer systems.

V. CONCLUSION AND FUTURE STUDY

Currently, telecommunications companies are going through a period of intense competition to retain current consumers by offering attractive new services (such as unlimited local and international calls, high-speed internet, and new phones) and lowering prices. Thus, it is critical to study and forecast customer churn behaviour with more precision. Analyzing churn data and developing a more accurate prediction model isn't easy due to the data's inherent imbalance. The suggested method successfully anticipates fraudulent phone calls by analyzing data from Call Detail Records (CDRs). When compared to the other models, XGBoost 96.7% accuracy was the best. The K-Means model has great promise for use in telecom fraud detection due to its high accuracy. Additional study has the potential to yield more effective outcomes in the field of telecommunications fraud detection, which encompasses a wider range of fraud detection strategies when evaluated side by side. Future work aims to improve precision using hybrid models, ensemble learning, and voice biometrics. Real-time deployment and testing on larger datasets will assess scalability. Deep learning and graph-based call analysis may further enhance performance.

REFERENCES

[1] I. Murynets, M. Zabarankin, R. P. Jover, and A. Panagia, "Analysis and detection of SIMbox fraud in mobility networks," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, IEEE, Apr. 2014, pp. 1519–1526. doi: 10.1109/INFOCOM.2014.6848087.

[2] U. Aslam, M. Jayabalan, H. Ilyas, and A. Suhail, "A survey on opinion spam detection methods," *Int. J. Sci. Technol. Res.*, vol. 8, no. 9, pp. 1355–1363, 2019.

[3] K. Mallikarjuna Rao Bhavikkumar Patel, "Suspicious Call Detection and Mitigation Using Conversational AI," *Defensive Publ. Ser.*, 2023.

[4] M. T. Aras and M. A. Guvensan, "Challenges and Key Points for Fraud Detection in Aviation," in *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, IEEE, Aug. 2021, pp. 1–6. doi: 10.1109/INISTA52262.2021.9548388.

[5] M. Arafat, A. Qusef, and G. Sammour, "Detection of Wangiri Telecommunication Fraud Using Ensemble Learning," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings*, 2019. doi: 10.1109/JEEIT.2019.8717528.

[6] M. Liu, J. Liao, J. Wang, and Q. Qi, "AGRM: Attention-Based Graph Representation Model for Telecom Fraud Detection," in *IEEE International Conference on Communications*, 2019. doi: 10.1109/ICC.2019.8761665.

[7] 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.

[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] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, Dec. 2015, doi: 10.1186/s40537-015-0029-9.

[10] 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.

[11] A. Marzuoli, H. A. Kingravi, D. Dewey, and R. Pienta, "Uncovering the landscape of fraud and spam in the telephony channel," in *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, 2017. doi: 10.1109/ICMLA.2016.195.

[12] P. Piyush, 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.

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

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

[15] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[16] M. Nyirenda and J. C. Daka, "Smart Mobile Telecommunication Network Fraud Detection System Using Call Traffic Pattern

Analysis and Artificial Neural Network," *Am. J. Intell. Syst.*, vol. 12, no. 2, pp. 43–50, 2023.

[17] V. R. Krishna and S. Boddu, "Financial Fraud Detection using Improved Artificial Humming Bird Algorithm with Modified Extreme Learning Machine," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 5s, pp. 05–14, May 2023, doi: 10.17762/ijritcc.v11i5s.6593.

[18] V. Prajapati, "Enhancing Threat Intelligence and Cyber Defense through Big Data Analytics : A Review Study," *J. Glob. Res. Math. Arch.*, vol. 12, no. 4, pp. 1–6, 2025.

[19] H. Kali, "The Future of HR Cybersecurity: AI-Enabled Anomaly Detection in Workday Security," *Int. J. Recent Technol. Sci. Manag.*, vol. 8, no. 6, pp. 80–88, 2023.

[20] K. Bhargavi and B. M. Shivani, "Detection of Fraudulent Phone Calls Detection in Mobile Applications," *Turkish J. Comput. Math. Educ.*, vol. 15, no. 2, pp. 1–5, May 2024, doi: 10.61841/turcomat.v15i2.14644.

[21] 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.

[22] G. Singh, P. Singh, and M. Singh, "Advanced Real-Time Fraud Detection Using RAG-Based LLMs," 2025.

[23] J. Zhao, Y. Huang, Q. Xie, W. Wang, L. Wang, and K. Wu, "Chameleon: An Adaptive System for Overlapping Keystroke Signal Separation and Identification," in *2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS)*, IEEE, Oct. 2024, pp. 60–67. doi: 10.1109/ICPADS63350.2024.00018.

[24] B. Hong, T. Connie, and M. K. O. 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.

[25] J. Zhang, H. Chen, X. Yao, and X. Fu, "CPFinder: Finding an unknown caller's profession from anonymized mobile phone data," *Digit. Commun. Networks*, vol. 8, no. 3, pp. 324–332, Jun. 2022, doi: 10.1016/j.dcan.2021.08.003.

[26] N. Kale, S. Kochrekar, R. Mote, and S. Dholay, "Classification of Fraud Calls by Intent Analysis of Call Transcripts," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2021, pp. 1–6. doi: 10.1109/ICCCNT51525.2021.9579632.

[27] S. M. Gowri, G. S. Ramana, M. S. Ranjani, and T. Tharani, "Detection of Telephony Spam and Scams using Recurrent Neural Network (RNN) Algorithm," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, Mar. 2021, pp. 1284–1288. doi: 10.1109/ICACCS51430.2021.9441982.

[28] V. Kavitha, G. H. Kumar, S. V. M. Kumar, and M. Harish, "Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms," *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 181–184, May 2020, doi: 10.17577/IJERTV9IS050022.

[29] L. Suhuan and H. Xiaojun, "Android Malware Detection Based on Logistic Regression and XGBoost," in *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, IEEE, Oct. 2019, pp. 528–532. doi: 10.1109/ICSESS47205.2019.9040851.

[30] M. Abdul Jabbar and Suharjito, "Fraud detection call detail record using machine learning in telecommunications company," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 4, pp. 63–69, 2020, doi: 10.25046/aj050409.

[31] J. Xing, M. Yu, S. Wang, Y. Zhang, and Y. Ding, "Automated Fraudulent Phone Call Recognition through Deep Learning," *Wirel. Commun. Mob. Comput.*, pp. 1–9, Aug. 2020, doi: 10.1155/2020/8853468.

[32] P. Gao, Z. Li, D. Zhou, and L. Zhang, "Reinforced Cost-Sensitive Graph Network for Detecting Fraud Leaders in Telecom Fraud," *IEEE Access*, vol. 12, pp. 173638–173646, 2024, doi: 10.1109/ACCESS.2024.3448260.

