Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 4, 2026, 487-500.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research Paper

A Hybrid AdaBoost-Random Forest Framework for Telecom Fraud Detection using CDRs

Zainab Hassan and Esmaeel Tahanian*

Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran.

Article Info

Article History:
Received 21 June 2026
Revised 14 July 2026
Accepted 06 August 2026

DOI:10.22044/jadm.2026.17863.2956

Keywords:
Telecommunications Fraud Detection, Ensemble Learning, Call Detail Record (CDR), AdaBoost, Random Forest, Imbalanced Data, SMOTE

**Corresponding author: e.tahanian@shahroodut.ac.ir (E. Tahanian).*

Abstract

Today, telecommunications fraud has emerged as a major challenge for operators, resulting in billions of dollars in financial losses annually. The presence of substantial noise and severe class imbalance between legitimate and fraudulent data complicates the identification of fraud patterns within massive volumes of Call Detail Records (CDRs). This paper proposes a hybrid ensemble model, termed Hybrid AdaBoost-RF, for telecommunication fraud detection. In this model, Random Forest is employed as the base learner within the AdaBoost framework to enhance the model's robustness against noise. Furthermore, the SMOTE technique is utilized to address the class imbalance problem. Additionally, we applied a decision threshold tuned on the training predictions to improve the model's sensitivity in detecting fraudulent behavior. Experimental results demonstrate that the proposed model outperformss existing methods in recent research, achieving a Recall of 0.87 and an F1-Score of 0.86 on the test partition of the evaluated CDR dataset using the adopted experimental protocol. The model also achieves a Recall of 0.87 and an F1-Score of 0.86. Moreover, the Area Under the Curve (AUC) for ROC and PR metrics reach 0.9777 and 0.8733, respectively, validating the high efficiency of the proposed model.

1. Introduction

Telecommunications fraud represents a growing challenge within the communications industry, inflicting substantial losses on network operators. According to statistics published by the CFCA 0 in 2023, global losses attributed to telecom fraud reached approximately $38.95 billion, marking a 12% increase compared to 2021. This figure accounts for nearly 2.5% of total telecommunications revenue. Beyond the staggering financial implications, these fraudulent activities undermine customer trust and lead to a degradation in the Quality of Service (QoS). Consequently, the development of robust mechanisms for detecting and preventing telecommunications fraud has become an urgent necessity.

Initially, many fraud detection systems relied on rule-based methods, which identified telecommunication fraud using predefined rule sets [2]. While effective for detecting known fraud types, these methods became obsolete as attackers devised more sophisticated techniques. Subsequently, extensive research has been conducted using Machine Learning (ML) and Deep Learning (DL) to address these limitations. Recent studies have focused on algorithms such as XGBoost and LightGBM 00.

However, these algorithms face significant challenges in operational environments characterized by substantial class imbalance. Additionally, the high level of noise inherent in Call Detail Records (CDR) can lead to an increased False Positive Rate (FPR), ultimately resulting in financial losses for operators. Consequently, the need for an efficient approach that can simultaneously manage data imbalance and

achieve an optimal trade-off between Recall and Precision remains a fundamental challenge.

In this research, to overcome the existing challenges in telecommunications fraud detection, a hybrid ensemble model, Hybrid AdaBoost-RF, was implemented, in which Random Forest serves as the base learner (weak learner) for the AdaBoost framework. The primary objective of this integration was to address the inherent sensitivity of standard Decision Trees to noisy CDR data and to enhance the model's overall detection accuracy. To manage the significant class imbalance between fraudulent and legitimate data, the SMOTE (Synthetic Minority Over-sampling Technique) was employed, which is one of the most widely recognized oversampling techniques. The performance of this approach was benchmarked against models used in previous studies, demonstrating the superiority of the proposed method. Evaluated on a dataset obtained from Kaggle, the model achieved an F1-Score of 86% and a Recall of 87%. To further validate the model's efficiency, ROC (Receiver Operating Characteristic) and Precision-Recall (PR) curves were analyzed, yielding Area Under the Curve (AUC) values of 0.9777 and 0.8733, respectively—results that signify a substantial improvement in detection performance.

The primary contributions and achievements of this research can be summarized as follows:

- Integration of AdaBoost with Random Forest as a robust base estimator: The primary novelty of the proposed framework lies in replacing the conventional decision stump used in AdaBoost with a shallow Random Forest. Unlike traditional AdaBoost implementations that rely on high-bias weak learners, the proposed framework exploits the bootstrap aggregation and feature subsampling mechanisms of Random Forest to improve robustness against noisy CDR data while preserving the adaptive reweighting strategy of AdaBoost. Consequently, the proposed Hybrid AdaBoost-RF framework provides a more stable boosting process for telecom fraud detection and better captures complex fraud patterns than conventional boosting approaches.
- Addressing Extreme Class Imbalance: The second contribution involves the strategic application of the SMOTE technique to resolve the extreme class imbalance problem, which often leads to a model bias toward the majority class. Implementing this technique maximized the Recall and, consequently, the F1-Score—a factor that is economically vital for operators to minimize financial losses.
- Optimal Decision Thresholding: The introduction of an optimized decision threshold for final predictions further refined the model's performance. This adjustment significantly increased the model's sensitivity in distinguishing fraudulent activities from legitimate calls.
- Comparative Validation: Finally, by re-implementing four state-of-the-art methods from recent literature on the current dataset, the decisive superiority of the proposed approach was empirically validated across all key performance metrics.

The remainder of this paper is organized as follows: Section 2 reviews the existing literature and relevant prior studies. Section 3 is dedicated to a detailed description of the proposed methodology and the techniques employed in this study. In Section 4, we provide an in-depth analysis of the dataset used, the primary evaluation metrics, and the rationale behind their selection. Subsequently, the results of the proposed model are comprehensively analyzed, and the model's sensitivity to various parameters and techniques is evaluated. Furthermore, four state-of-the-art methods from recent literature are re-implemented on the current dataset to discuss their respective strengths and weaknesses in comparison to the Hybrid AdaBoost-RF model. Finally, Section 5 presents the conclusions and suggests directions for future work.

2. Related Works

Fraud detection across various domains has emerged as a dynamic and evolving field within machine learning research. Historically, investigations have transitioned from traditional, rule-based systems toward more sophisticated machine learning (ML) and deep learning (DL) paradigms, with increasingly advanced techniques developed to counteract fraudulent activities. The following review of prior studies is organized into three distinct subsections: traditional and base learning approaches, strategies employed to mitigate data imbalance, and the latest advanced hybrid methodologies.

2.1. Fraud Detection Systems and Approaches

This section focuses on general methodologies employed for fraud detection across diverse

domains. In a study conducted by Jain et al. 0 in 2019, various machine learning techniques were evaluated and compared, with a specific emphasis on their inherent limitations. Their research assessed a wide array of algorithms, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), Bayesian Networks, Fuzzy Logic-Based Systems, K-Nearest Neighbors (KNN), Hidden Markov Models, and Decision Trees [4]. The findings indicated that while some of these methods demonstrate relatively strong performance in controlled experimental environments, they face a critical challenge: there is no guarantee that an algorithm yielding acceptable results on one dataset will maintain the same level of efficacy on another. Ultimately, Jain et al. concluded that transitioning toward ensemble models is essential to ensure robustness and generalizability across varying data distributions.

In another study conducted by Min and Lin (2018) ,[٥] the K-Means algorithm—an unsupervised learning approach—was investigated for fraud detection within telecommunications signaling data. To enhance the model's efficiency, they employed Principal Component Analysis (PCA) for dimensionality reduction and utilized Grid Search to optimize clustering parameters.0

Their findings revealed that while this methodology is effective for identifying initial suspicious clusters, it exhibits significant sensitivity to the initialization of cluster centroids and the inherent noise present in telecommunication data. Consequently, the need for a more robust technique capable of mitigating such noise further underscores the importance of our proposed model. In another study focusing on telecommunications network monitoring, Shaeiri et al. [24] introduced a behavior-based online anomaly detection framework utilizing Apache Spark to identify malicious patterns within a nationwide Short Message Service (SMS) network. Their approach highlights the effectiveness of user profiling and statistical risk modeling in handling large-scale behavioral data. However, while their method mainly focuses on SMS activity traces, detecting complex fraud schemes within Call Detail Records (CDRs) presents distinct challenges—specifically severe class imbalance and noise in tabular feature sets—which our proposed Hybrid AdaBoost-RF framework directly addresses.

2.2. Handling Imbalanced Data in Machine Learning

In 2024, Zhu et al. 0 investigated an approach based on Neural Networks (NN) combined with the SMOTE technique. Their study addressed the inherent challenge of significant class imbalance in credit card transactions. By implementing SMOTE, they demonstrated that substantial improvements in Recall, Precision, and F1-Score reflect the model's robustness in managing the minority class.0

However, it is noteworthy that NN-based models typically require substantially larger training datasets compared to other methodologies. Furthermore, they can exhibit higher sensitivity to noise than tree-based methods, a factor that limits their efficacy in environments with highly volatile data such as CDRs.

In a similar vein, Mohammed et al. (2020) 0 conducted a comprehensive study comparing the efficacy of oversampling and undersampling techniques. Their investigation spanned a diverse set of machine learning algorithms, including Decision Trees, Random Forest, AdaBoost, Logistic Regression, and Support Vector Machines (SVM), among others. The findings of Mohammed et al. indicated that oversampling techniques consistently outperform undersampling methods, yielding superior scores across key performance metrics such as Recall, Precision, and F1-Score, as well as the ROC-AUC score.0

2.3. Ensemble Learning and Hybrid Architectures

In a study conducted by Randhawa et al. (2018) 0, machine learning algorithms were utilized for credit card fraud detection. Their research initially evaluated base models, followed by ensemble techniques, including AdaBoost and Majority Voting. To assess performance, they employed a real-world dataset from a financial institution. Furthermore, to rigorously test the model's reliability, controlled noise was intentionally introduced into the data. The results clearly demonstrated that the Majority Voting approach achieved superior accuracy compared to other methods. Ultimately, the findings provide strong empirical evidence for the increased efficacy of ensemble methodologies over individual base classifiers.

3. Proposed Method

This section presents the proposed framework for fraud detection in telecommunication networks. The primary objective of this model is to enhance the detection accuracy of fraudulent activities by distinguishing them from legitimate calls through the analysis of CDR 0.

Figure 1 illustrates the proposed methodology employed in this study. The process begins with the

ingestion of raw data and proceeds through sequential stages of preprocessing and data imbalance management before being fed into the proposed Hybrid AdaBoost-RF model.

The core innovation of this approach lies in utilizing Random Forest (RF) as the base (weak) learner, replacing the shallow decision trees typically used in standard AdaBoost. This modification is specifically designed to address the challenges associated with noise and the inherent complexity of telecommunication data. In the following subsections, each phase of this algorithm will be examined in detail.

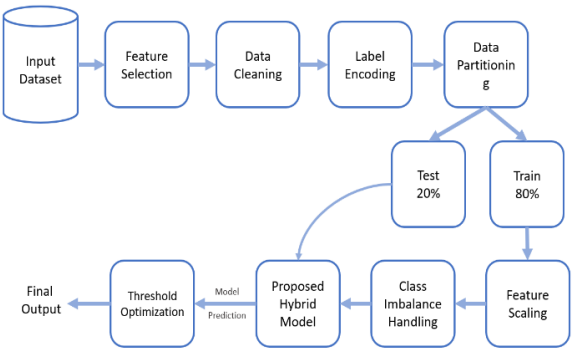

Figure 1. Overview of the proposed model process.

3.1. Data Preparation

The process begins with the acquisition of raw CDR data, followed by the removal of redundant columns, such as phone numbers, which are non-essential for the modeling process. Although these fields serve as unique identifiers for individual subscribers, they possess no direct predictive value for the model. Furthermore, such features exhibit high cardinality, which can inadvertently lead to overfitting in final predictions and increase the computational complexity without enhancing the model's discriminative power.

In the subsequent stage, the label column—which contains Boolean values representing either fraudulent or legitimate calls—is processed. To ensure compatibility with the learning algorithms, these categorical values are encoded into a binary format, where 0 represents legitimate calls and 1 denotes fraudulent activity. This column is designated as the target variable and is accordingly separated from the feature set to prepare the data for the training phase.

Subsequently, we addressed the issue of missing values, which is a common occurrence in telecommunications datasets. An initial Data Integrity Check was performed to evaluate the completeness of the dataset. The results of this analysis revealed that the specific dataset utilized in this research contained no missing values; therefore, no further imputation or missing data handling techniques were required. This ensures that the original distribution of the data remains intact for the subsequent modeling phases.

Following the preprocessing phase, the dataset was partitioned into training and testing sets using an 80:20 ratio. Given the extreme class imbalance between fraudulent and legitimate transactions inherent in telecommunications data, the application of stratified sampling was essential. This technique ensures that the minority class distribution remains consistent across both the training and testing subsets, thereby preventing biased evaluations and ensuring that the model is tested on a representative sample of the overall data distribution.

Following the data partitioning, feature scaling was performed. This stage is critical because the features in our dataset (such as day_min and init_calls) exhibit diverse ranges and units. Many algorithms that rely on gradient-based computations or iterative optimization, including AdaBoost, are sensitive to the scale of input features.

To address this, the StandardScaler technique was employed 0. This method rescales the features to ensure they have a mean of zero and a unit variance of one. This transformation is applied to each data point x for every feature using the following formula [10]:

$$z = \frac{(x - \mu)}{\sigma} \tag{1}$$

where μ represents the sample mean of the training set, and σ denotes the standard deviation of the training set. To strictly prevent data leakage, these parameters (mean and standard deviation) were computed exclusively from the training data and subsequently applied to transform the test data. This approach ensures that no information from the test set is utilized during the model's training phase, thereby maintaining the integrity and generalizability of the experimental results.

3.2. Data Augmentation via SMOTE

Following the feature scaling phase, we addressed the issue of class imbalance to prevent the model from developing a bias toward the majority class (legitimate calls). To establish a balance between legitimate and fraudulent instances, the SMOTE (Synthetic Minority Over-sampling Technique) was applied exclusively to the training data 00.

SMOTE is widely regarded as one of the most effective strategies for mitigating class imbalance. Rather than merely replicating existing instances, it enriches the feature space of the minority class by synthesizing new samples through the K-Nearest

Neighbors (KNN) algorithm. Consequently, this technique enhances the model's decision boundary, making it more robust in identifying fraudulent activities. The impact of the SMOTE application on the training data distribution is illustrated in Figure 2.

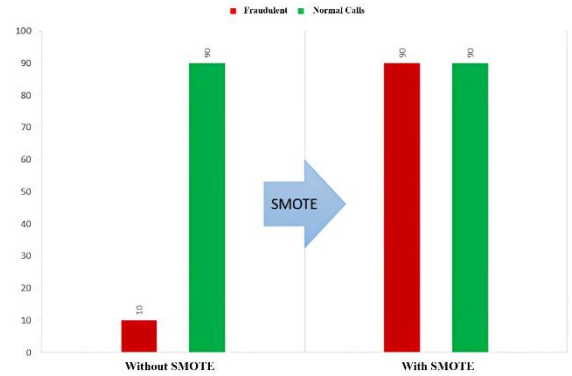

Figure 2. SMOTE operational methodology.

3.3. The Proposed Hybrid Ensemble Architecture

In this research, we employed AdaBoost, an ensemble learning technique that transforms a collection of weak learners into a single robust predictor through an iterative process 0.

The core innovation of our proposed Hybrid AdaBoost-RF model lies in the selection of the base learner. In standard AdaBoost implementations, shallow Decision Trees (often referred to as decision stumps) with high bias are typically utilized as weak learners 0.

However, to overcome the limitations of such models, we integrated the Random Forest (RF) algorithm—a significantly more powerful and robust ensemble method—as the base learner within the AdaBoost framework. This integration is designed to leverage the strength of RF in handling complex data patterns while benefiting from the adaptive boosting mechanism. The architecture of the proposed model is illustrated in Figure 3[14].

The Random Forest (RF) algorithm operates by initially selecting multiple distinct subsets of the CDR data through a process of random sampling with replacement, known as bootstrapping, which are then used to train individual trees. To ensure de-correlation among these trees, a random subset of features is selected at each node to determine the optimal split. This mechanism significantly reduces the overall variance of the final model. Consequently, Random Forest exhibits a substantially lower variance compared to a shallow Decision Tree.

This inherent characteristic makes our proposed hybrid model significantly more robust and stable when dealing with the high levels of noise typical in telecommunications data.

Figure 3. Architecture of the proposed model.

Furthermore, considering that CDR data in telecommunication transactions often encompasses numerous features and high dimensionality, the adoption of more advanced models like RF—instead of traditional weak learners—is instrumental in enhancing the overall performance. In practice, Random Forest leverages feature subsampling, whereby only a random subset of features is selected for training at each step. This mechanism effectively reduces inter-tree correlation, thereby significantly augmenting the model's generalization power and mitigating the risks associated with the "curse of dimensionality." Although conventional AdaBoost is traditionally constructed using shallow decision trees (decision stumps) as weak learners, such estimators often exhibit high bias and become highly sensitive to noisy observations in telecom CDRs. This behavior causes AdaBoost to progressively increase the weights of mislabeled or noisy samples, potentially leading to overfitting during later boosting iterations.

To alleviate this limitation, the proposed framework replaces the conventional weak learner with a shallow Random Forest. Owing to its bootstrap aggregation and random feature selection mechanisms, Random Forest produces more stable predictions while preserving sufficient diversity across boosting iterations. Consequently, AdaBoost focuses its reweighting process on genuinely difficult fraud patterns rather than random noise.

This structural integration distinguishes the proposed Hybrid AdaBoost-RF framework from conventional boosting approaches that rely solely on individual decision trees.

In this model, we utilized a depth of 6 for the Random Forest. This value was determined through hyperparameter tuning across lower depths. The rationale for maintaining a shallow

depth for the trees in Random Forest is to prevent over-complexity of the base learner and to preserve the incremental learning characteristic of AdaBoost.

In this structure, the learning process was conducted over 200 iterations[1]. This implies that at each stage, AdaBoost analyzes the error of the previous RF model and updates the weights of misclassified instances—particularly complex fraudulent data located near the normal patterns—according to the following equation:

$$w_{t+1} = \frac{w_{t,i} \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \quad (2)$$

In this equation, $h_t(x_i)$ denotes the output of the Random Forest model at iteration t. In the event of a misclassification, the weight of that instance is increased exponentially, ensuring that the RF model assigns higher priority to it in the subsequent iteration. This mechanism enables the RF to filter out noise while allowing AdaBoost to refine the decision boundary for fraudulent data more precisely.

3.4. Optimization and Decision Logic

Finally, to prioritize fraud detection (high Recall) while controlling false positives (reasonable Precision), a decision threshold τ is applied to the model's output probability scores for final binary classification.

This threshold was determined by maximizing the F1-Score on the Precision-Recall curve, using the model's predictions exclusively on the training set (after applying SMOTE balancing and completing the full training of the Hybrid AdaBoost-RF model).

By performing threshold tuning solely on training data, we ensure that no information from the test set influences the decision rule, thereby preventing any test set leakage or optimistic bias in the evaluation.

4. Experiment

This section is dedicated to examining and discussing the data used, the results obtained from the Hybrid AdaBoost-RF model, and the rationale for its superiority over previously proposed models. For model training, Python version 3.12.6 was employed. Additionally, Pandas was used to read the input dataset, Scikit-learn for modeling, and Imblearn for the data augmentation phase and SMOTE implementation. The summary of the execution steps for this model is presented in the Algorithm 1.

Algorithm 1: Hybrid AdaBoost-RF for Telecom Fraud Detection.

Input: Raw CDR Dataset D, Number of iterations T=200, Learning rate η = 0.4
Output: H(x) Final Classification Model

1. Step 1: Data Pre-processing
   - Remove non-predictive features (e.g., Phone Number).
   - Encode target labels: $y_i \in \{-1, 1\}$
   - Split data into D_{train} and D_{test} (80/20 ratio) using Stratified Sampling.
   - Standardize features: $z = (x-\mu)/\sigma$ using StandardScaler
2. Data Augmentation Phase:
   - Apply SMOTE on D_{train} to balance class distribution.
3. Hybrid Ensemble Training Phase:
   - Initialize sample weights: $w_{1,i} = 1/N$ for i=1,…,N
   - For t=1 to T:
     - Train Random Forest (base learner h_t) with max_depth=6 using w_t
     - Calculate error: $\epsilon_t = \sum_{i=1}^{N} w_{t,i} . \mathbb{I}(h_t(x_i \neq y_i)$
     - Compute estimator weight: $\alpha_t = \eta . \ln(\frac{1-\epsilon_t}{\epsilon_t})$
     - Update sample weights: $w_{t+1,i} = \frac{w_{t,i} . \exp(-\alpha_t . y . h_t(x_i))}{z_t}$
     - Normalize w_{t+1} such that $\sum w_{t+1,i} = 1$
   - End for
4. Inference & Decision Phase:
   - Aggregate weak learners: $H_{score}(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$
   - Determine optimal threshold τ by maximizing F1-Score on Precision-Recall curve.
   - Final Decision: $Y_{pred} = 1$ if $H_{score} \geq \tau$ else 0

4.1. Dataset Description

The data used in this paper were extracted from Kaggle, titled "Fraud Detection using Call Detail Records"[15].

This dataset comprises 17 feature columns and 101,174 instances of subscriber activities. A detailed description of each column is provided in Table 1.

The most significant characteristic of fraud detection datasets is the presence of class imbalance. Based on the analysis conducted on this dataset, the ratio of legitimate to fraudulent instances was determined as Table 2.

This Table demonstrates that a severe class imbalance exists within our dataset. In traditional methods, such an imbalanced distribution causes the model to develop a bias toward the majority class. As previously described, we addressed this issue by employing the SMOTE technique.

Table 1. Dataset features for model training.

Feature	Data Type	Description

Phone Number	Nominal/ID	Customer's unique phone identifier.
Account Length	Integer	Total duration of the subscriber's membership.
VMail Message	Integer	Total number of voicemail messages.
Day Mins	Float	Total duration of calls made during the day.
Day Calls	Integer	Total number of calls made during the day.
Day Charge	Float	Total cost of calls made during the day.
Eve Mins	Float	Total duration of calls made during the evening.
Eve Calls	Integer	Total number of calls made during the evening.
Eve Charge	Float	Total cost of calls made during the evening.
Night Mins	Float	Total duration of calls made during the night.
Night Calls	Integer	Total number of calls made during the night.
Night Charge	Float	Total cost of calls made during the night.
Intl Mins	Float	Total duration of international calls.
Intl Calls	Integer	Total number of international calls.
Intl Charge	Float	Total cost of international calls.
CustServ Calls	Integer	Total number of calls made to customer service.
isFraud	Boolean	Target variable: fraudulent or legitimate activity.

Table 2. Distribution ratio of normal and fraudulent data.

Class	Number of Samples	Distribution Percentage
0 (Non-Fraud)	90,642	89.59%
1 (Fraud)	10,532	10.41%
Total	101,174	100%

4.2. Performance Results

The results described in this section were obtained from the evaluation of the model on the test data, which comprised 20% of the dataset. The model had not encountered these data during the training phase. In this research, accuracy was not used to evaluate the model's performance, as this metric can be highly misleading in the context of imbalanced data. Consider a model that classifies all calls as legitimate; such a model would achieve an accuracy of over 90% simply because it correctly identifies the majority of normal calls, despite failing to detect any instances of fraud. Consequently, our primary evaluation metrics are as follows:

- Precision: Represents the proportion of all cases identified as fraud by the model that were actually fraudulent [16].
  where TP, TN, FP, and FN denote True Positives, True Negatives, False Positives, and False Negatives, respectively.

$$\Pr ecision = \frac{TP}{TP + FP} \quad (3)$$

- Recall: Represents the proportion of all actual fraudulent cases that were correctly identified by the model.

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \quad (4)$$

- F1-Score: The harmonic mean of precision and recall.

$$F1 - Score = \frac{\Pr ecision \times \mathrm{Re}\,call}{\Pr ecision + \mathrm{Re}\,call} \quad (5)$$

- ROC-AUC Curve: A numerical value between 0 and 1 that indicates the model's ability to distinguish between classes across all thresholds 0.
- PR-AUC Curve: Directly reflects the model's performance on the minority class (fraud) 0.

The confusion matrix illustrates the model's detection performance on the test data. This has been a standard metric for evaluating fraud detection models in similar domains; therefore, we examine the confusion matrix in this research [19]. As depicted in Figure 4, the model successfully identified the majority of fraudulent instances (True Positives). This directly contributes to the increase in the model's recall, where a higher value signifies a substantial reduction in financial losses. Furthermore, the very low rate of false negatives in this matrix—representing fraudulent cases incorrectly identified as legitimate—demonstrates the high level of security the model provides for telecommunication operators.

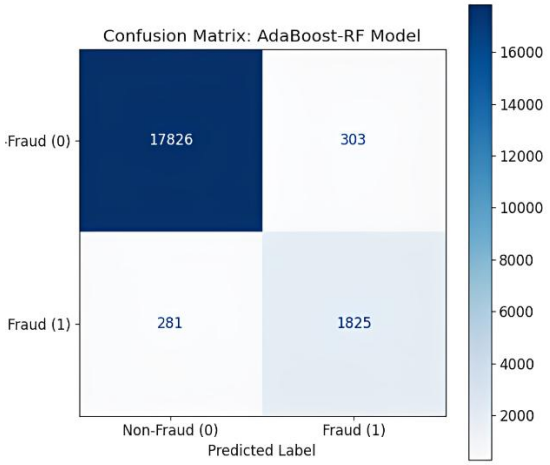

Figure 4. Confusion Matrix for the proposed model.

In Figure 5, the classification report is presented, providing a comprehensive overview of the model's performance on both normal and fraudulent call categories. In this report, the values for precision, recall, and F1-score are calculated approximately.

	precision	recall	f1-score	support
No Fraud (0)	0.98	0.98	0.98	18129
Fraud (1)	0.86	0.87	0.86	2106
accuracy			0.97	20235
macro avg	0.92	0.92	0.92	20235
weighted avg	0.97	0.97	0.97	20235

Figure 5. Classification Report for the proposed model.

Based on the data provided in Figure 5, the model's performance in detecting normal data is exceptionally strong. Furthermore, the high recall value in the fraud class confirms the effectiveness of the SMOTE and AdaBoost techniques in identifying fraudulent instances.

Additionally, the high F1-score indicates that the model has successfully established an effective balance between recall and precision. Consequently, it can be concluded that the hybrid AdaBoost-RF model, integrated with the SMOTE technique, provides a highly effective approach for fraud detection in CDR data.

The PR-AUC curve is one of the most critical evaluation metrics for datasets with severe class imbalance. By plotting the trade-off between precision and recall across all possible thresholds, this curve provides a precise assessment of the model's performance specifically on the minority class. Figure 6 illustrates the curve obtained from the proposed model:

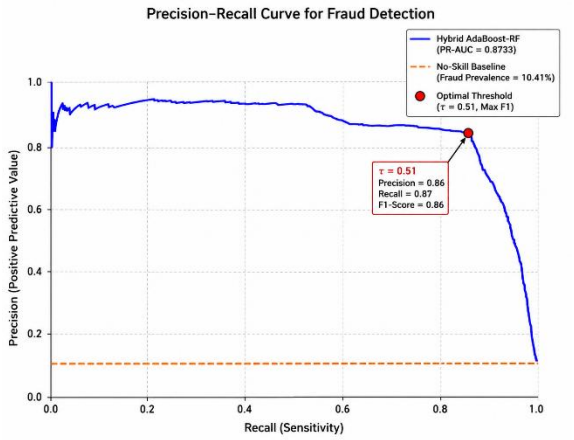

Figure 6. Precision–Recall (PR) curve of the proposed Hybrid AdaBoost-RF model evaluated on the held-out test set. The blue curve represents the precision–recall trade-off across all classification thresholds (PR-AUC = 0.8733). The dashed orange line denotes the no-skill baseline corresponding to the fraud prevalence (10.41%), while the red marker indicates the optimal decision threshold (τ = 0.51).

As you can observe, the area under the precision-recall curve (PR-AUC) is 0.8733, indicating strong model performance. This value demonstrates the extent to which the model can maintain its precision while achieving a high recall. The closer this value is to 1.0, the more successful the model has been in reaching this objective.

The next metric we employ to evaluate the model is the area under the ROC curve (ROC-AUC). The closer this value is to 1 (indicating perfect performance), the higher the model's capability. In essence, this value indicates the probability that the model will correctly classify a truly fraudulent instance. Figure 7 illustrates the curve generated by the model.

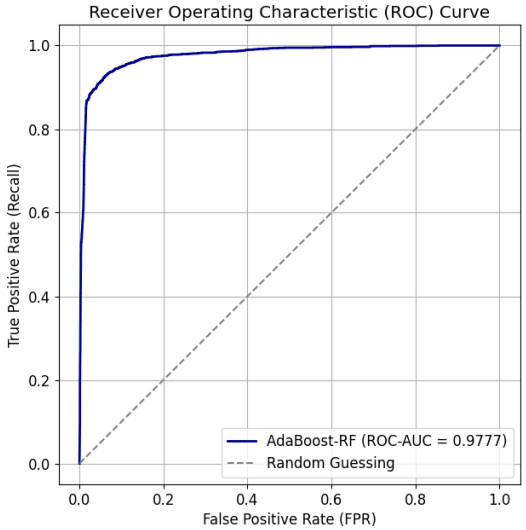

Figure 7. ROC-AUC achieved for the proposed model.

The area under the curve (AUC) is 0.9777, which is very close to 1.0, demonstrating the model's outstanding performance.

The ROC-AUC and PR-AUC curves presented in this section were generated using the predictions obtained on the held-out test partition of the CDR dataset.

4.3. Ablation Study and Comparative Analysis

To gain a deeper understanding of the model's performance and ensure it is properly tuned, this section examines the impact of various components, such as feature scaling methods and different hyperparameter values.

All ablation experiments reported in this study were conducted using the same train-test split and identical experimental settings, with only the evaluated hyperparameter varied in each experiment. The results presented correspond to a single experimental run under this fixed evaluation protocol to ensure a fair comparison among the investigated configurations.

4.3.1. Data Balancing Impact

The use of Random Forest (RF), due to its tree-based structure and sub-sampling, can partially

address the class imbalance issue. However, as will be demonstrated, this alone is insufficient. Implementing balancing techniques separately can significantly improve the model's performance.

To identify the most effective data balancing method, we evaluated the model in three scenarios:

1. Without balancing
2. With SMOTE (Synthetic Minority Over-sampling Technique)
3. With ADASYN (Adaptive Synthetic Sampling),[20] [21].

The reason for excluding undersampling techniques from this experiment was to avoid the loss of a substantial amount of data, which could be critical for the model's training and performance.

In this approach, all parameters used across the three scenarios were kept identical, with only the balancing technique being varied. The results obtained for the fraudulent data are presented in the Table 3:

Table 3. Comparison of resampling results in the proposed model.

Technique	Recall	Precision	F1-Score
None	0.84	0.85	0.85
SMOTE	0.87	0.85	0.86
ADASYN	0.87	0.84	0.85

As you can observe in Table 3, transitioning from the unbalanced state to the balancing techniques resulted in a 3% improvement in Recall. This increase is highly significant because a higher Recall value directly correlates with reduced financial losses for operators, as fewer fraudulent activities go undetected.

Between the SMOTE and ADASYN techniques, the SMOTE method achieved higher Precision and, consequently, a better F1-Score. This indicates that it provides the optimal balance between Recall and Precision for our model.

Therefore, it can be concluded that the application of balancing methods can significantly contribute to reducing financial losses for telecommunications operators. Furthermore, the SMOTE technique was selected as the superior approach for strengthening decision boundaries, as it establishes an optimal balance between Recall and Precision.

4.3.2. Hyperparameter Sensitivity Analysis

To ensure reproducibility, the final hyperparameter configuration was determined through an empirical sensitivity analysis. Rather than adopting default parameter values, several candidate configurations were evaluated while keeping all remaining parameters fixed. The learning rate was fixed at 0.4 after preliminary experiments demonstrated a stable balance between convergence speed and robustness. Subsequently, different values of the Random Forest maximum depth and the number of AdaBoost estimators were systematically evaluated based on Recall, F1-Score, and PR-AUC. The final configuration was selected as the one providing the best overall trade-off between fraud detection performance and model complexity.

Unless otherwise specified, all hyperparameters remained fixed during the sensitivity analysis. Specifically, the learning rate was set to 0.4, the Random Forest criterion was Gini impurity, bootstrap sampling remained enabled, and only the hyperparameter under investigation (either max_depth or n_estimators) was varied in each experiment.

The performance of the proposed model is highly dependent on the optimal tuning of its hyperparameters. In this section, we examine the impact of two key parameters: max_depth (the maximum depth of the trees) and n_estimators (the number of boosting iterations). Our primary criteria for selecting the best values are maximizing the F1-Score and Recall, while simultaneously preventing excessive model complexity and overfitting.

Determining Sensitivity of max_depth (Tree Depth in Random Forest)

In this section, we trained the model using depths of 4, 5, and 6. These values were chosen to maintain model simplicity while utilizing an effective method to achieve optimal results. In all these experiments, to ensure the integrity and consistency of the results, we kept all other parameters constant and varied only the tree depth. In Table 4, you can observe the results obtained from this comparison:

Table 4. Comparison of results obtained from different RF depths.

Max_depth	Recall	Precision	F1-Score
4	0.82	0.84	0.83
5	0.87	0.85	0.86
6	0.87	0.86	0.86

Based on the experimental results, a maximum tree depth of 6 was selected because it achieved the highest overall performance while maintaining a relatively simple model structure. Increasing the depth beyond this value was considered unnecessary, as deeper trees would increase model complexity without providing a meaningful improvement in the evaluated performance metrics.

As observed in Table 4, increasing the depth from 4 to 5 leads to a significant increase in Recall, rising from 82% to 87%. This improvement is vital for

operators as it directly enhances the detection of fraudulent activities.

Furthermore, increasing the depth from 5 to 6 resulted in an approximately 1% increase in Precision.

For a more comprehensive analysis and to decide between choosing a depth of 5 or 6, other critical metrics such as the PR-AUC curve were examined. Figure 8 illustrates a performance comparison of all three models using this curve:

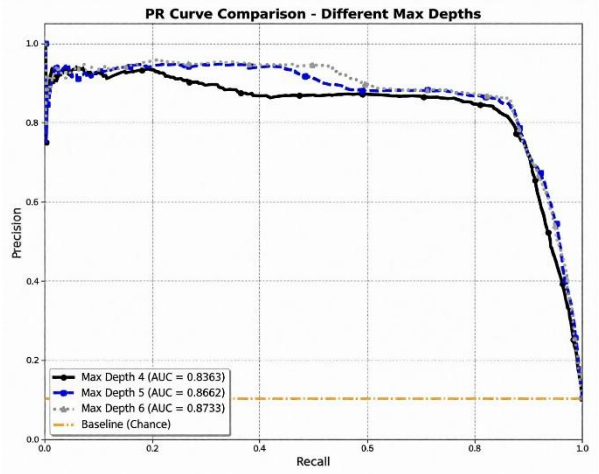

Figure 8. Comparison of different depths for Random Forest.

As illustrated in chart 8, the model with a depth of 6 demonstrates superior performance in fraud detection across all possible thresholds compared to the models with depths of 4 and 5.

Therefore, it can be concluded that to maintain the optimal balance between Recall and Precision, and to achieve the best possible performance on the fraud class, a depth of 6 is the ideal value for the trees. It is worth noting that despite the improved results at depth 6, we avoided further increasing the tree depth to prevent overfitting on the noisy CDR data and to maintain the incremental learning nature of the AdaBoost model.

Determining Sensitivity of n_estimators

After fixing the value of max_depth, we examined the values of 100, 150, and 200 for n_estimators and their impact on the model's results. In this section, similar to the previous part, we kept other parameters constant and identical for all three experiments, varying only the n_estimators value.

In Table 5, we have provided a comprehensive comparison of the results obtained from all three experiments:

Table 5. Comparison of results from different n_estimators values.

n_estimator	100	150	200
Recall	0.88	0.87	0.87
Precision	0.84	0.85	0.86
F1-Score	0.86	0.86	0.86
PR-AUC	0.8641	0.8699	0.8733
ROC-AUC	0.9766	0.9774	0.9777

Similarly, 200 boosting iterations were selected because they provided the best overall balance between Recall, F1-Score, and PR-AUC among the evaluated configurations. Although additional boosting iterations may further increase computational cost, the observed performance improvements beyond 200 estimators were expected to be marginal. Therefore, 200 estimators were adopted as the final configuration.

The model with n_estimators=100 achieved the highest Recall but the lowest Precision. By increasing n_estimators at each stage, the values for Precision, PR-AUC, and ROC-AUC also increased. Upon reaching 200 iterations, the model achieved the best performance among all iterations and reached better stability.

Although increasing the number of iterations from 150 to 200 improved the model's results, and it can be observed that the results continue to improve as iterations increase, increasing the number of iterations beyond this leads to an increase in computational load without significant improvement; therefore, the value n_estimators=200 was selected for the final model.

Comparative Analysis

In order to accurately evaluate the method proposed in this paper, we selected previous works in similar fields and implemented them on our dataset. This implementation allows us to assess our model's performance against similar methods and demonstrate its superiority.

- In the method proposed by Taha et al. [22] in 2020, they suggested a method named OLightGBM, which was an Optimized LightGBM approach that utilized Information Gain for feature selection and Bayesian Optimization for parameter tuning. We implemented the method proposed in that paper on our dataset [15] and reached the results shown in Table 6.

  As is evident in Table 6, this model demonstrated very good performance on our dataset, although the achieved Recall was 4% lower than our proposed model, which is relatively significant and could impose substantial losses on telecommunications operators. Additionally, it achieved an equal Precision, but its F1-Score was also lower than our model due to the lower Recall.
- The next paper, published by Arafat et al. [3] in 2019, focuses on detecting Wangiri fraud in the telecommunications sector. They utilized

the XGBoost Classifier algorithm and changed the decision threshold from the default value of 0.5 to 0.2. The objective of this approach was to manage the extreme data imbalance; consequently, this adjustment made the model more sensitive in detecting fraud. Inspired by the method used in Arafat's work, we implemented this algorithm on our dataset and achieved the results shown in Table 7 for the fraud class.

As observed, although this method achieved a high Recall of 0.71, it was simultaneously accompanied by a sharp drop in Precision, which consequently resulted in an F1-Score=0.64.

In our proposed method, in addition to determining the threshold, we use SMOTE to manage imbalanced data, which prevents the decrease in Precision and provides better results.

- The third research study reviewed and re-implemented was by Wu et al. [23] in 2018. This paper, using the Artificial Immune System (AIS) as the weak learner for AdaBoost, was based on laboratory telecommunications data. Based on the results obtained from implementing this method on our dataset, we achieved the results shown in Table 7 [23].

  As you can observe in Table 7, despite achieving a very high Recall (0.91), we obtained a very low Precision and, consequently, a very low F1-Score. The most important reason for these results was that Wu et al. used a balanced laboratory dataset; as a result, the AIS model identifies anything that slightly deviates from their normal data as fraud.

  In our proposed method, by using an imbalanced dataset and balancing it with oversampling techniques, as well as using RF as the base model, we overcame these problems.

- In the final assessment conducted, we used the traditional AdaBoost method, which utilizes a Decision Tree with low depth as the weak learner for AdaBoost. This method has been employed in many papers. [14] To ensure the superiority of the RF model over Decision Tree (DT), we tested tree depths from 1 to 6 to prove that despite the increased complexity of the DT model, our proposed method still maintains its superiority. The results obtained from these experiments can be seen in Table 6.

Table 6. Comparison of results from max_depth in standard AdaBoost.

Max_depth	F1-Score	Recall	Precision
1	0.28	0.53	0.19
2	0.30	0.58	0.20
3	0.32	0.55	0.23
4	0.35	0.52	0.27
5	0.43	0.55	0.35
6	0.50	0.56	0.45

As you can observe in Table 6, with the increase in tree depth, the results improve; however, they still have a significant difference compared to our proposed method. To ensure a fair comparison, all baseline models were re-implemented using the same dataset and identical experimental protocol adopted for the proposed Hybrid AdaBoost-RF framework. Specifically, the same 80:20 stratified train-test split, preprocessing pipeline, feature standardization procedure, and evaluation metrics were consistently applied across all compared models. Whenever applicable, the same data balancing strategy (SMOTE) was also employed to minimize experimental bias and ensure that performance differences were attributable to the classification models rather than differences in the evaluation protocol.

Finally, in Table 7, a summary comparison of the performance of the Hybrid AdaBoost-RF model with the methods used in other papers is provided. As you can observe, our proposed model has the highest results among all the suggested algorithms:

Table 7. Summary of results from related works.

Method	Reference	Recall	Precision	F1-Score
OLightGBM	Taha (2020)	0.83	0.86	0.84
XGBoost	Arafat (2019)	0.71	0.58	0.64
AdaBoost + AIS	Wu (2018)	0.91	0.14	0.24
AdaBoost + DT	Gedela (2022)	0.56	0.45	0.50
Proposed Hybrid	This work	0.87	0.86	0.86

4.4. Practical Deployment Considerations

The proposed Hybrid AdaBoost-RF framework is designed to support deployment in practical telecom fraud detection environments through a two-stage architecture. During the offline stage, the model is trained using historical labeled CDR data. Since model training involves multiple boosting iterations and Random Forest construction, this process can be executed periodically on dedicated computing resources without interfering with operational network services.

Once deployed, the framework performs only inference on incoming CDRs. Because prediction requires only evaluating the trained ensemble, the

computational cost during operation is considerably lower than that of the training stage, making the proposed approach suitable for near-real-time fraud screening.

As fraud patterns evolve over time, the model can be periodically retrained using newly collected labeled data to maintain detection performance under changing fraud behaviors.

Furthermore, the proposed framework can be integrated into existing telecom fraud management systems as a decision-support component that assigns fraud probabilities to incoming transactions before billing or customer management processes are executed.

4.5. Limitations of the Proposed Framework

Although the proposed Hybrid AdaBoost-RF framework demonstrates strong performance on the evaluated telecom fraud dataset, several limitations should be acknowledged. First, the model is trained and evaluated using a single labeled CDR dataset. Consequently, its performance may vary when applied to datasets with different feature distributions, fraud characteristics, or customer behavior patterns. Second, the proposed framework relies on supervised learning and therefore requires sufficiently labeled historical fraud data for effective training. In environments where labeled fraud instances are scarce, model performance may be affected. Furthermore, although SMOTE effectively alleviates class imbalance during training, synthetic samples may not fully represent newly emerging fraud strategies or rapidly evolving fraud patterns.

As fraud behaviors change over time, periodic retraining using recently collected labeled data is recommended to maintain detection performance. Future work will investigate adaptive learning techniques capable of handling concept drift and improving the generalization ability of the proposed framework across heterogeneous telecom environments.

We acknowledge that the experimental evaluation reports point estimates of the performance metrics without statistical significance testing or confidence intervals. Although the reported results were obtained using a fixed experimental protocol, future work will include repeated experiments with different random seeds and statistical hypothesis testing to provide a more comprehensive assessment of the proposed framework.

4.6. Model Interpretability and Decision Support

Model interpretability is an important consideration for practical fraud detection systems, where security analysts often require explanations for the generated predictions. Since the proposed framework is based on tree-based ensemble models, it naturally supports global feature importance analysis, enabling operators to identify the variables that contribute most significantly to fraud detection. Such information can help telecom analysts understand the general characteristics of fraudulent activities and prioritize network monitoring efforts. In addition, the probabilistic output generated by the Hybrid AdaBoost-RF model allows suspicious transactions to be ranked according to their estimated fraud risk, enabling operators to allocate investigation resources more efficiently.

Furthermore, the proposed framework is compatible with widely used post-hoc explainability techniques, such as SHAP and LIME, which can provide instance-level explanations without modifying the underlying classification model. These capabilities enhance the practical usefulness of the proposed framework as a decision-support tool for telecom fraud management.

5. Conclusion

In this paper, a Hybrid AdaBoost-RF framework is proposed for telecommunications fraud detection using CDR data. The model uses Random Forest as the base learner within AdaBoost to handle noise and improve minority class detection stability. Additionally, SMOTE addresses class imbalance, while decision threshold tuning enhances detection sensitivity.

Experimental results show that the proposed Hybrid AdaBoost-RF outperforms baseline models across all key metrics, achieving Recall=0.87, F1-Score=0.86, ROC-AUC=0.9777, and PR-AUC=0.8733. These results demonstrate the model's efficiency in mitigating telecommunications fraud and reducing associated financial losses for operators.

In future work, we will investigate more advanced hybrid methods, including the use of Deep Learning approaches such as Autoencoders, as well as working on more robust ensemble models that combine several algorithms like XGBoost, CatBoost, and LightGBM as an ensemble to leverage the advanced features of these models.

References

[1] V. Banu, "Real-time fraud detection in telecom charging systems using AI," *International Journal of Emerging Trends in Computer Science and Information Technology*, pp. 571-582, 2025.

[2] R. Afzal and R. K. Murugesan, "Implementation of a malicious traffic filter using snort and wireshark as a proof of concept to enhance mobile network security," *Journal of Telecommunications and Information Technology, no. 1*, pp. 64-71, 2022.

[3] A. Balouchi, M. Abdollahi, A. Eskandarian, K. Karimi Pour Kerman, E. Majd, N. Azouji, and A. Baniasadi, "Wangiri fraud detection: A comprehensive approach to unlabeled telecom data," *Future Internet, vol. 18, no. 1*, p. 15, 2025.

[4] Y. Jain, N. Tiwari, S. Dubey, and S. Jain, "A comparative analysis of various credit card fraud detection techniques," *International Journal of Recent Technology and Engineering, vol. 7, no. 5*, pp. 402-407, 2019.

[5] X. Min and R. Lin, "K-means algorithm: fraud detection based on signaling data," in *2018 IEEE World Congress on Services (SERVICES)*, 2018, pp. 21-22.

[6] M. Zhu, Y. Zhang, Y. Gong, C. Xu, and Y. Xiang, "Enhancing credit card fraud detection: a neural network and SMOTE integrated approach," *arXiv preprint arXiv:*2405.00026, 2024.

[7] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th International Conference on Information and Communication Systems (ICICS*), 2020, pp. 243-248.

[8] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access, vol.* 6, pp. 14277-14284, 2018.

[9] E. A. Desta, K. W. Azale, A. A. Hailu, F. B. Adugna, A. T. Mengesha, S. F. Belay, et al., "Enhancing subscription fraud detection through ensemble learning: the case of Ethio telecom," *Scientific Reports*, 2026.

[10] Z. L. Thakker and S. H. Buch, "Effect of feature scaling pre-processing techniques on machine learning algorithms to predict particulate matter concentration for Gandhinagar, Gujarat, India," *International Journal of Scientific Research in Science and Technology*, vol. 11, no. 1, pp. 410-419, 2024.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research,* vol. 16, pp. 321-357, 2002.

[12] P. Sundaravadivel, R. A. Isaac, D. Elangovan, D. KrishnaRaj, V. L. Rahul, and R. Raja, "Optimizing credit card fraud detection with random forests and SMOTE," *Scientific Reports*, vol. 15, no. 1, p. 17851, 2025.

[13] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[14] B. Gedela and P. R. Karthikeyan, "Credit card fraud detection using AdaBoost algorithm in comparison with various machine learning algorithms to measure accuracy, sensitivity, specificity, precision and F-score," in *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, 2022, pp. 1-6.

[15] J. Furgoson, "Fraud detection using call detail records," Kaggle Dataset, 2024. [Online]. Available: https://www.kaggle.com/datasets/jakefurgoson/fraud-detection-using-call-detail-records

[16] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *European Conference on Information Retrieval, Berlin, Heidelberg*, 2005, pp. 345-359.

[17] M. Altalhan, A. Algarni, and M. T. H. Alouane, "Imbalanced data problem in machine learning: A review," *IEEE Access*, 2025.

[18] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 233-240.

[19] A. Saputra, "Fraud detection using machine learning in e-commerce," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, 2019.

[20] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322-1328.

[21] J. Brandt and E. Lanzén, "A comparative review of SMOTE and ADASYN in imbalanced data classification," *2021*.

[22] A. A. Taha and S. J. Malebary, "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine," *IEEE Access*, vol. 8, pp. 25579-25587, 2020.

[23] B. Wu, M. Li, and C. Zhou, "Application of AdaBoost algorithm and immune algorithm in telecommunication fraud detection," in *2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018),* 2018, pp. 159-163.

[24] Z. Shaeiri, J. Kazemitabar, S. Bijani, and M. Talebi, "Behavior-based online anomaly detection for a nationwide short message service," *Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 2, pp. 319-331, 2019.

یک چارچوب ترکیبی AdaBoost-جنگل تصادفی برای کشف تقلب در مخابرات با استفاده از CDR ها

زینب حسن و اسماعیل طحانیان*

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شاهرود، شاهرود، ایران.

ارسال ۲۰۲۶/۰۶/۱۱؛ بازنگری ۲۰۲۶/۰۷/۱۴؛ پذیرش ۲۰۲۶/۰۸/۰۶

چکیده:

امروزه، تقلب در صنعت مخابرات به چالش بزرگی برای اپراتورها تبدیل شده و سالانه میلیاردها دلار خسارت مالی به همراه دارد. وجود نویز زیاد و عدم توازن شدید کلاسی میان داده‌های قانونی و تقلب‌آمیز، شناسایی الگوهای تقلب را در حجم عظیم داده‌های جزئیات تماس پیچیده می‌سازد. این مقاله یک مدل یادگیری جمعی ترکیبی به نام Hybrid AdaBoost-RF را برای کشف تقلب در مخابرات پیشنهاد می‌کند. در این مدل، جنگل تصادفی به عنوان یادگیرنده پایه در چارچوب AdaBoost به کار گرفته شده است تا مقاوم‌سازی مدل در برابر نویز افزایش یابد. علاوه بر این، از تکنیک SMOTE برای رفع مشکل عدم توازن کلاس‌ها استفاده شده است. همچنین، جهت بهبود حساسیت مدل در شناسایی رفتارهای تقلب‌آمیز، یک آستانه تصمیم‌گیری تنظیم‌شده بر اساس پیش‌بینی‌های آموزش اعمال گردید. نتایج تجربی نشان می‌دهد که مدل پیشنهادی با استفاده از پروتکل تجربی اتخاذشده، از روش‌های موجود در پژوهش‌های اخیر عملکرد بهتری داشته و در بخش داده‌های آزمون مجموعه داده CDR ارزیابی‌شده، به بازخوانی (Recall) ۰/۸۷ و F1-Score برابر با ۰/۸۶ دست یافته است. همچنین، سطح زیر منحنی برای معیارهای ROC و PR به ترتیب به ۰/۹۷۷۷ و ۰/۸۷۳۳ رسیده است که کارایی بالای مدل پیشنهادی را تأیید می‌کند.

کلمات کلیدی: کشف تقلب در مخابرات، یادگیری ترکیبی، داده‌های جزئیات تماس، جنگل تصادفی، AdaBoost، داده‌های نامتوازن.