ACTA INFOLOGICA 2023;7(1):125-138
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1222890

RESEARCH ARTICLE

An Intrusion Detection Approach based on the Combination of Oversampling and Undersampling Algorithms

Örneklem Arttırma ve Örneklem Azaltma Algoritmalarının Kombinasyonuna Dayalı Bir Saldırı Tespit Yaklaşımı

Ahmet Okan Arık[1], Gülsüm Çiğdem Çavdaroğlu[2]

ABSTRACT

The threat of network intrusion has become much more severe due to the increasing network flow. Therefore, network intrusion detection is one of the most concerned areas of network security. As demand for cybersecurity assurance increases, the requirement for intrusion detection systems to meet current threats is also growing. However, network-based intrusion detection systems have several shortcomings due to the structure of the systems, the nature of the network data, and uncertainty related to future data. The imbalanced class problem is also crucial since it significantly negatively affects classification performance. Although high performance has been achieved in deep learning-based methodologies in recent years, machine learning techniques may also provide high performance in network intrusion detection. This study suggests a new intrusion detection system called ROGONG-IDS (Robust Gradient Boosting - Intrusion Detection System) which has a unique two-stage resampling model to solve the imbalanced class problem that produces high accuracy on the UNSW-NB15 dataset using machine learning techniques. ROGONG-IDS is based on gradient boosting. The system uses Synthetic Minority Over-Sampling Technique (SMOTE) and NearMiss-1 methods to handle the imbalanced class problem. The proposed model's performance on multi-class classification was tested with the UNSW-NB15, and then its robust structure was validated with the NSL-KDD dataset. ROGONG-IDS reached the highest attack detection rate and F1 score in the literature, with a 97.30% detection rate and 97.65% F1 score using the UNSW-NB15 dataset. ROGONG-IDS provides a robust, efficient intrusion detection system for the UNSW-NB15 dataset, which suffered from imbalanced class distribution. The proposed methodology outperforms state-of-the-art and intrusion detection methods.

Keywords: Machine learning, cyber security, intrusion detection system, imbalanced data, gradient boosting

ÖZ

Artan ağ akışı nedeniyle ağa izinsiz giriş tehdidi çok daha şiddetli hale gelmiştir. Bu nedenle, ağ güvenliğinde en çok endişe duyulan alanlardan biri ağ saldırı tespitidir. Siber güvenlik güvencesine olan talep arttıkça mevcut tehditleri karşılamak için saldırı tespit sistemlerine olan gereksinim de artmaktadır. Bununla birlikte, ağ tabanlı saldırı tespit sistemlerinin, sistemlerin yapısı, ağ verilerinin doğası ve gelecekteki verilerle ilgili belirsizlik nedeniyle bazı eksiklikleri vardır. Dengesiz veri problemi de sınıflandırma performansını kötü etkilediği için çok önemlidir. Son yıllarda derin öğrenme tabanlı metodolojilerde yüksek performans elde edilmesine rağmen, makine öğrenme teknikleri de ağ saldırı tespitinde yüksek performans sağlayabilir. Bu çalışma, makine öğrenme tekniklerini kullanarak UNSW-NB15 veri setinde yüksek doğruluk üreten dengesiz sınıf problemini çözmek için benzersiz bir iki aşamalı yeniden örnekleme modeline sahip olan ROGONG-IDS (Robust Gradient Boosting - Saldırı Tespit sistemi) adlı yeni bir saldırı tespit sistemi önermektedir. ROGONG-IDS, gradyan artırmaya dayalıdır. Sistem, dengesiz sınıf problemini çözmek için Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) ve NearMiss-1 yöntemlerini kullanır. Önerilen modelin çok sınıflı sınıflandırma performansı UNSW-NB15 ile test edilmiş, güçlü yapısı NSL-KDD veri seti ile doğrulanmıştır. ROGONG-IDS, UNSW-NB15 veri setini kullanarak %97,30 tespit oranı ve %97,65 F1 skoru ile literatürdeki en yüksek saldırı tespit oranı ve F1 skoruna ulaşmıştır. ROGONG-IDS, dengesiz sınıf dağılımından muzdarip UNSW-NB15 veri kümesi için sağlam, verimli bir saldırı tespit sistemi sağlamaktadır. Önerilen metodoloji, literatürdeki en gelişmiş saldırı tespit metotlarından daha iyi performans göstermektedir.

Anahtar Kelimeler: Makine öğrenmesi, siber güvenlik, saldırı tespit sistemi, dengesiz veri, gradyan artırma

[1](PhD), Istanbul University, Istanbul, Turkiye
[2](Assist. Prof. Dr.), Isik University, Faculty of Economics, Administrative and Social Sciences, Department of Information Technologies

ORCID: A.O.A. 0000-0002-6572-1605; G.Ç.Ç. 0000-0002-4875-4800

Corresponding author:
Ahmet Okan ARIK
Istanbul University, Istanbul, Turkiye
E-mail address: aokanarik@gmail.com

Submitted: 23.12.2022
Revision Requested: 17.02.2023
Last Revision Received: 14.02.2023
Accepted: 17.04.2023
Published Online: 14.06.2023

Citation: Arik, A.O., Cavdaroglu, G.C. (2023). An intrusion detection approach based on the combination of oversampling and undersampling algorithms. *Acta Infologica, 7*(1), 125-138. https://doi.org/10.26650/acin.1222890

1. INTRODUCTION

According to the Mobility Report of Ericsson, mobile network data traffic grew 42 percent between Q3 2020 and Q3 2021. Total monthly mobile network data traffic in Q3 2021 reached around 78EB (Ericsson, 2021). Due to stay-at-home activities, the Covid-19 pandemic resulted in a spike in network traffic from 2019-2020. This increase also varied regionally. The most rapid growth of international internet bandwidth was experienced in Africa, growing at a compound annual rate of 45% between 2017 and 2021. A 38% compound annual rate during the same period was experienced in Oceania (Mauldin, 2021). Hence, the threat of network intrusion has become much more severe. Consequently, network intrusion detection is considered one of the significant concerns in the network domain. As the demand for cybersecurity assurance increases, the requirement for intrusion detection systems (IDS) to meet current threats is also growing. IDS can be divided into three groups according to the collection mechanisms: (1) Network-based IDS (NIDS), (2) Host-based IDS (HIDS), and (3) Hybrid IDS. The main goal of a HIDS is to monitor network traffics in a particular host and analyze the file system, login activities, and currently running processes. On the other hand, NIDS detects any attacks on the hosts of that network. Hybrid IDS models use both of them.

NIDS has several shortcomings due to the structure of the systems, the nature of the network data, and uncertainty related to future data. First, NIDS schemes are occasionally insufficient since they can detect normal/abnormal attacks, but not the exact attack type. Secondly, the up-to-datedness of data sets in which the schemes are tested is crucial because detecting emerging attacks is necessary to develop a scheme that will work with high performance in modern networks. However, most NIDS schemes were tested on outdated data sets. Finally, the imbalanced class problem is also crucial since it has a significant effect on classification performance (Zhang, Huang, Wu, & Li, 2020). Imbalanced network intrusion data makes it hard to detect minority attack classes accurately.

According to the detection techniques used in these systems, IDS can be categorized as (1) Misuse-based IDS and (2) Anomaly-based IDS. Anomaly-based IDS have increasingly attracted attention in recent years since the other types of IDS can identify only known attacks and suffer from the incompetence to detect new attacks. This study proposes a unique two-stage resampling method within the scope of ROGONG-IDS, developed to detect minority class attacks that anomaly-based IDS types have difficulty detecting.

The techniques used in anomaly-based IDS can be divided into (1) Machine Learning methods and (2) Deep Learning methods. *Fig. a* shows these methods and their sub-methods.

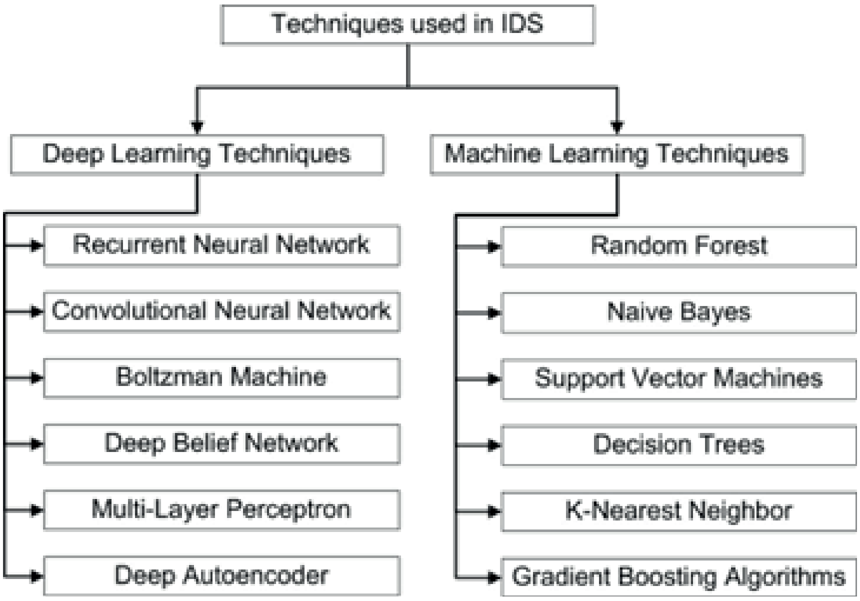


Figure a. Techniques used in anomaly-based IDS

Although high performance has been achieved in methods developed using deep learning techniques in recent years, models that produce results with high accuracy (ACC) in wide-ranging data sets can be developed using machine learning techniques. It was indicated that machine learning techniques, such as support vector machine (SVM), random forest (RF), and decision tree (DT), are insufficient to distinguish normal and abnormal network activities due to the diversification of attack categories and the surge of network traffic (Zhang *et al*., 2020). However, it was observed that the gradient boosting technique used in the presented study produced results with high performance in the UNSW-NB15 dataset, which has a dramatically imbalanced class problem. Consequently, this study presents a new method that produces results with high ACC on imbalanced data sets using machine learning techniques. The presented method currently has the highest performance in the literature.

The contributions of the ROGONG-IDS (Robust Gradient Boosting - Intrusion Detection System) method are as follows:

1. Shortening of testing times: Since ROGONG-IDS provides a highly effective and low-complexity model for the feature selection and classifier method, testing times are shortened. Due to its short testing time, ROGONG-IDS is suitable for real-time network environments.

2. A two-step solution method for the imbalanced class problem: ROGONGIDS offers a unique, robust, and effective two-step solution for the imbalanced class problem.

3. The ROGONG-IDS method has the highest ACC (97.30%) and F_1 score (97.65%) in the literature.

4. The ROGONG-IDS method can be helpful in frequently incorporating gradient boosting methods and resampling studies into IDS development, which are rarely used in IDS models. Therefore, the method triggers the development of IDS in this direction.

The rest of the paper is organized as follows. Section 2 discusses the related work of machine learning and deep learning techniques used in Anomaly-based IDS. Section 3 provides a brief description of the ROGONG-IDS method. Experiment results are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes the work.

2. RELATED WORK

The techniques used in IDS can be examined under two headings according to the methods used. First-generation techniques were created using Machine Learning methods, and second-generation techniques were created using Deep Learning methods. This section will discuss the techniques available in the literature, and performance measures will be examined.

2.1. Machine learning techniques

Chkirbene, Eltanbouly, Bashendy, AlNaimi, & Erbad (2020) proposed a hybrid approach that combines two machine learning algorithms. The proposed methodology detects possible attacks by performing effective feature selection and classification. They used the RF algorithm to find the essential features, classification, and regression trees (CART) to classify the different attack classes. They tested the approach using the UNSW-NB15 dataset. The ACC is 95.73% for the UNSW-NB15 dataset and 97.03% for the KDD99 dataset.

Injadat, Moubayed, Nassif, & Shami (2021) proposed a novel multi-stage optimized machine learning-based framework to reduce computational complexity and maintain detection performance. They evaluated the framework's performance using the CICIDS 2017 and the UNSW-NB15 datasets. The detection accuracies are over 99% for both datasets.

Bhavani, Rao, & Reddy (2020) proposed a mix of the DT and RF algorithms. The ACC of the DT algorithm is 81.86%, and the F_1 score is 82.00%. On the other hand, the ACC of the RF algorithm is 95.32%, and the F_1 score is 95.00%. According to the results, the RF algorithm is a better method to overcome the over-fitting problem

Kaja, Shaout, & Ma (2019) proposed a two-stage architecture. The architecture based on machine learning algorithms uses K-Means to detect attacks in the first stage; it uses supervised learning to classify such attacks and eliminate the number of false positives. The ACC of their approach is 99.95%, and the F_1 score is 99.99%.

Belouch, El Hadaj, & Idhammad (2018) presented a study that evaluates the performances of several machine learning methods, such as SVM, Naive Bayes (NB), DT, and RF. The evaluation criteria used in this study are ACC, building time, and prediction time. They used the UNSW-NB15 dataset to evaluate the mentioned methods. The detection rate (DR), also known as recall, and precision of the RF method considered the best method according to this study, are 97.49% and 93.53%, respectively.

2.2. Deep learning techniques

Sumaiya Thaseen, Saira Banu, Lavanya, Rukunuddin Ghalib, & Abhishek (2021) proposed a new methodology based on deep learning techniques. The methodology includes a correlation-based feature selection phase integrated with neural network for identifying anomalies. They tested the approach on the UNSW-NB15 and KDD99 datasets. The performance results show that their approach is superior in ACC, sensitivity, and specificity compared to some state-of-the-art techniques. The overall ACC is 96.44% for the UNSW-NB15 dataset.

The model proposed by Liu, Gao, & Hu (2021) addresses the imbalanced data problem. An ensemble model is used to solve the imbalanced data problem in the presented study. This model uses ADASYN for oversampling and LightGBM for classification. After normalization, the authors evaluated the model using the KDD, UNSW-NB15, and CICIDS2017 datasets. The model, which offers a more prosperous and shorter training time than other IDS models, reached 85.89% ACC on the UNSW-NB15 dataset.

Mulyanto, Faisal, Prakosa, & Leu (2021) proposed a cost-sensitive neural network based on focal loss (FL-NIDS) to overcome the imbalanced data problem. To evaluate the UNSW-NB15, NSL-KDD, and Bot-IoT intrusion detection datasets, they applied this system using a convolutional neural network (CNN). The ACC score does not reflect the DR of the minority classes. They evaluated the approach using the F_1 score. For the UNSW-NB15 dataset, the CNN-SMOTE model reached a 36% F_1 score, while FL-NIDS reached a 39% F_1 score.

Zhang *et al.* (2020) proposed a flow-based IDS model. They developed a new class imbalance processing technology for large-scale data and combined it with CNN. Their methodology's DR, precision, and F_1 score on the UNSW-NB15 dataset were 96.54%, 98.30%, and 97.26%, respectively. This study is currently the one that produces results with the highest performance in the literature.

Andresini, Appice, Mauro, Loglisci, & Malerba (2020) proposed a multi-channel deep learning method called MINDFUL. It combines an unsupervised approach with a supervised one. The unsupervised one is for multi-channel feature construction. This phase is based on two encoder neural networks. The supervised approach is for exploiting cross-channel feature correlations. They have tested the method on the KDDCUP99Test, UNSW-NB15Test, and CICIDS2017Test datasets. The performance criterion for the UNSW-NB15 dataset is 93.40% ACC and 95.29% F_1 score.

Khan, Gumaei, Derhab, & Hussain (2019) proposed a two-stage IDS model. The model first detects whether the network packets are normal or abnormal based on the probability score generated by the stacked auto-encoder (AE). The attacks are then classified using the Softmax classifier. Therefore, the model can also classify unlabeled data. The model, which was evaluated with different algorithms, reached 89.13% ACC, 0.74 false alarm rate (FAR) on the UNSW-NB15 dataset.

Yang, Zheng, Wu, & Yang (2019) combined an improved conditional variational AE with a DNN. The trained encoder was used to automatically reduce data dimension and initialize the weight of DNN hidden layers. This way, the DNN can quickly achieve global optimization through backpropagation and fine-tuning. They used the NSL-KDD and UNSW-NB15 datasets to evaluate the performance of their model. According to the results, the proposed method shows better performance metrics than the nine intrusion detection methods. The ACC, DR, precision, F_1 score, and FPR metrics are 85.97%, 77.43%, 97.39%, 86.27%, and 2.74, respectively.

Zhang *et al.* (2018) proposed a new network intrusion detection scheme based on deep learning techniques. Reducing the feature dimensionality is crucial for network intrusion systems. Therefore, the proposed scheme used a denoising auto-encoder (DAE) with a weighted loss function for feature selection. The selected data is classified using a compact multi-layer perceptron

(MLP) to identify intrusions. The proposed scheme was tested on the UNSW-NB15 dataset. The DR, precision, and F_1 score of the proposed scheme on the UNSW-NB15 dataset were 98.80%, 95.98%, and 95.2%, respectively.

Mulyanto *et al.* (2021) utilized deep neural networks (DNNs) to predict the attacks on network IDS. They compared the approach with Ada Boost, DT, K-Nearest neighborhood (K-NN), linear regression, NB, RF, and SVM methods. As a result, one-layer DNN architecture achieved 92.9% ACC and 95.4% F_1 score and outperformed traditional machine learning techniques.

Naseer *et al.* (2018) developed anomaly detection models based on different DNN structures, including CNN, AEs, and recurrent neural networks (RNNs). They trained the models on the NSL-KDD training dataset and evaluated them using the NSL-KDDTest+ and NSL-KDDTest21 datasets. According to the study results, CNN reached 85% ACC, and LSTM reached 89% ACC.

Yin, Zhu, Fei, & He (2017) proposed a deep learning approach for intrusion detection using RNNs (RNN-IDS). They studied the model's performance in binary and multi-class classification and compared it with J48, artificial neural network (ANN), RF, SVM, and other machine learning methods. They tested the model on the NSL-KDD dataset. Their model reached 81.29% ACC with RNN and 78.10% ACC with MLP.

Table I

Briefly provides some of the advanced anomaly-based IDS models' performances on the UNSW-NB15 dataset in terms of development and testing

Author	Method	Classification Type	Accuracy (%)	Detection Rate (%)	F_1 Score (%)	Precision (%)
Chkirbene *et al.* (2020)	ML	Multiclass	95.73	-	-	-
Injadat *et al.* (2021)	ML	Binary	99	-	-	-
Belouch *et al.* (2018)	ML	Binary	97.49	93.53	-	-
Sumaiya Thaseen *et al.* (2021)	DL	Multiclass	96.44	-	-	-
Zhang *et al.* (2020)	DL	Multiclass	96.54	96.54	98.30	97.26
Andresini et al. (2020)	DL	Binary	93.40	-	95.29	-
Khan *et al.* (2019)	DL	Multiclass	89.13	-	-	-
Yang *et al.* (2019)	DL	Multiclass	85.97	77.43	86.27	97.39
Zhang *et al.* (2018)	DL	Binary	-	98.80	95.2	95.98
Mulyanto *et al.* (2021)	DL	Multiclass	92.9	95.4	-	-

Table i. Advanced anomaly-based IDS models use the UNSW-NB15 dataset in the literature.
(ML: machine learning, DL: deep learning)

3. METHOD

The ROGONG-IDS method consists of three modules, as shown in *Fig. b*. Data has been made suitable for modeling with the operations performed in the data preprocessing module. The processing of categorical data is provided with one-hot encoding and label encoding applied in this module. Feature selection was made to improve ACC by reducing the training time of the model and removing redundant features. Data standardization has been applied to examine different types of measurable features in a common standard. The module that handles the imbalanced class problem includes resampling operations to ensure class balance. The resampling method has two stages: the NearMiss-1 method for undersampling and the SMOTE method for oversampling. The resampling method is the most critical method that increases the model's ACC. Finally, several gradient boosting method tests to decide the classification decision method. Hyperparameters of the selected method are optimized using Bayesian optimization.

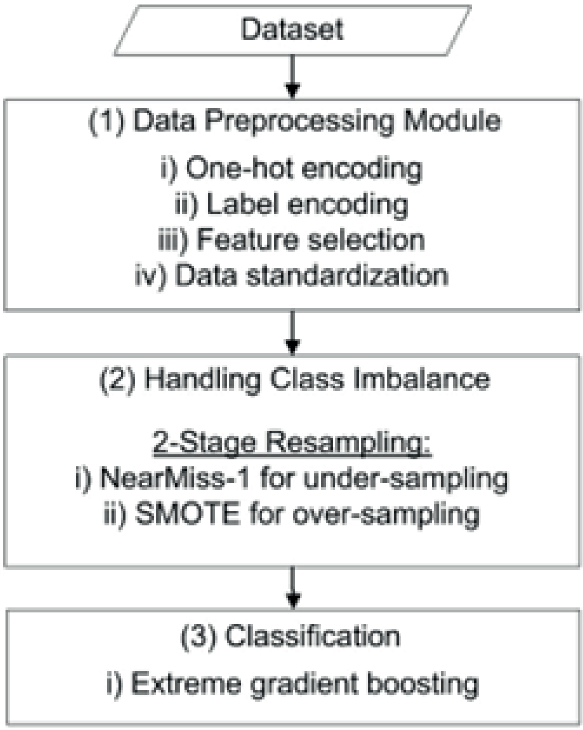

Dataset
(1) Data Preprocessing Module
i) One-hot encoding
ii) Label encoding
iii) Feature selection
iv) Data standardization
(2) Handling Class Imbalance
2-Stage Resampling:
i) NearMiss-1 for under-sampling
ii) SMOTE for over-sampling
(3) Classification
i) Extreme gradient boosting

Figure b. Flow of the proposed method

3.1.1. Data preprocessing

The data preprocessing module was carried out feature selection, one-hot encoding, label encoding, and data standardization processes. The DAE developed from Zhang *et al.* (2018), which reduces the feature size by limiting the number of critical features, was used for feature selection. This process selected twelve properties: Dtcpb, Stcpb, Service, Dload, Dmeansz, Service dns, Smeansz, Sload, Trans depth, Sttl, Service ftp-data, and Ct ftp. The UNSW-NB15 dataset has three nominal properties:" proto"," state", and" service". These attributes have 135, 16, and 14 different values, respectively. In order to process these features with machine learning algorithms, the one-hot encoding technique is used while maintaining the irregular relationship. With this process, the data set's features increased from 47 to 208. In addition, label encoding was applied to the target feature attack class. Finally, features were standardized to ensure that the estimators equally weighted numerical features of different units and scales. Z Score Scaling is used for standardization in the study. This method sets the mean value to 0 and the standard deviation equal to 1 for each feature, making the data have comparable scales.

3.1.2. Handling imbalanced class problem

The classes that account for a significant part of the data set are named majority classes, and the classes that account for the minor are named minority classes. The data set of skewed class balances is called the imbalanced data set. This is because classifiers are highly sensitive to the majority and less sensitive to the minority classes. Imbalanced classes need to be balanced to develop high-accuracy intrusion, detection models. The most typical strategy for balancing class distributions is using different resampling methods (Haibo He & Yunqian Ma, 2013). These are undersampling and oversampling methods. Undersampling aims to reduce the number of majority class observations at a specified rate or number. On the other hand, oversampling aims to increase the number of observations of minority classes at a determined rate or number (Chawla, Bowyer, Hall, & Kegelmeyer, 2002; Haibo He & Yunqian Ma, 2013).

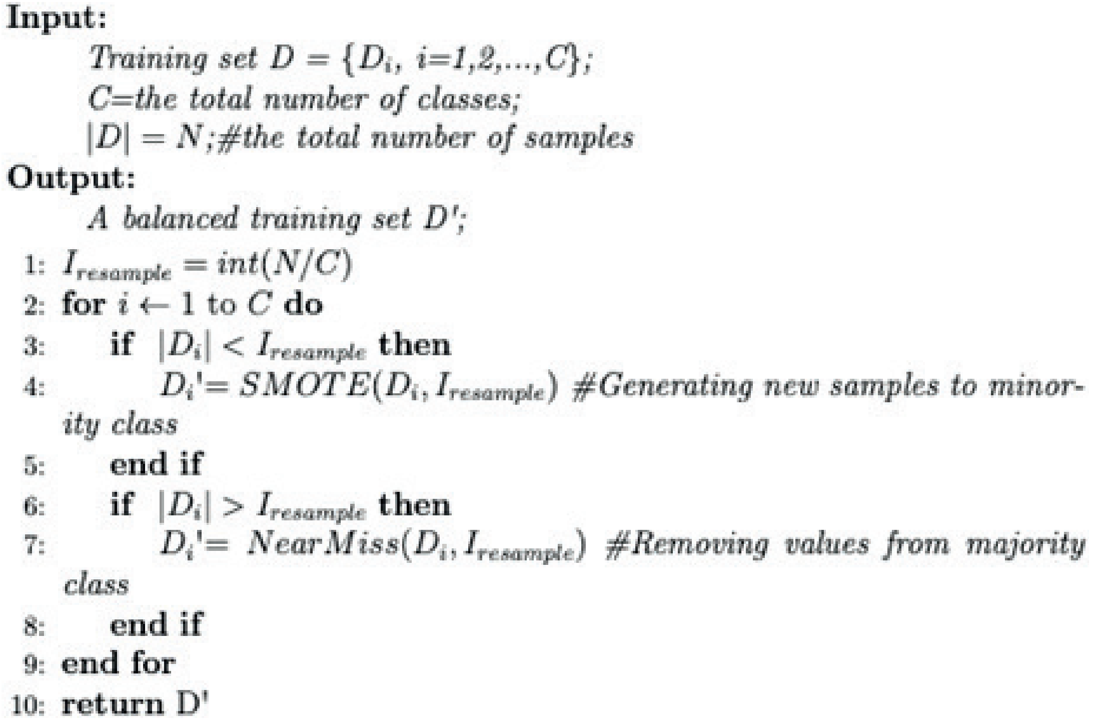
The UNSW-NB15 dataset contains highly imbalanced class distribution. Among the 2,000,054 million samples, there are two classes with less than 2,000 instances. Undersampling methods cause information loss by reducing the number of observations in the majority class, thus reducing the representativeness of classes (Demidova & Klyueva, 2017). Studies indicate that oversampling methods work better than undersampling methods in handling the problem of imbalanced data because they do not cause data loss. However, there are also disadvantages to applying only oversampling methods while achieving class balance. It can significantly increase the data, thus increasing the computational cost. Another disadvantage is that it may lead to an overfitting problem. Accordingly, oversampling for these minority classes or undersampling for majority classes is insufficient to solve the imbalanced data problem. Therefore, the ROGOND-IDS model uses a method that recommends using oversampling and undersampling methods together to overcome this challenge. As shown in *Table II*, 10 different undersampling methods were tried to find the proper one that produced the most successful result. The NearMiss-1 undersampling method produced the best results with oversampling method SMOTE. NearMiss-1 (Yen & Lee, 2006) selects majority class observations close to some minority class observations. Class balance is achieved by calculating the minimum distance between the majority class observation and the three nearest minority class instances in this undersampling method. On the other hand, SMOTE (Chawla *et al*., 2002) is an oversampling method used in generating minority class samples. It generates a new observation based on the similarity of a minority class instance and its nearest neighbor. Depending on the amount of oversampling sample required, neighbors are selected from the k nearest neighbors using the k-NN algorithm. This means that observations similar to existing minority class examples are generated (Demidova & Klyueva, 2017). These samples generated with SMOTE prevent overfitting and improve the classifier's performance.

Using this 2-stage resampling method, the data set is resampled with all classes containing an equal number of samples $\boldsymbol{Iresample = int} \, (\frac{\boldsymbol{number\ of\ samples}}{\boldsymbol{number\ of\ classes}})$. In this way, imbalanced class distribution is prevented.

Table ii
Studied undersampling methods (ACC: accuracy, DR: detection rate).

Undersampling Method	Accuracy (%)	F_1 Score (%)	Detection Rate (%)	Algorithm	Oversampling Method
AIIKNN	96.05	96.77	96.05	XGBoost	SMOTE
Edited Nearest Neighbours	96.26	96.89	96.26	XGBoost	SMOTE
Repeated Edited Nearest Neighbours	96.08	96.78	96.08	XGBoost	SMOTE
Instance Hardness Threshold	94.20	95.00	94.20	XGBoost	SMOTE
NearMiss (v1)	96.49	97.10	96.49	XGBoost	SMOTE
NearMiss (v3)	95.45	96.41	95.45	XGBoost	SMOTE
Neighbourhood Cleaning Rule	96.03	96.64	96.03	XGBoost	SMOTE
Random Undersampling	96.22	96.97	96.22	XGBoost	SMOTE
Tomek Link	96.24	96.89	96.24	XGBoost	SMOTE
One Sided Selection	95.44	96.41	95.44	XGBoost	SMOTE

Algorithm 1 shows the pseudocode of ROGONG-IDS, which handles the imbalanced class problem.



```
Input:
    Training set $D = \{D_i,\ i=1,2,\ldots,C\}$;
    C=the total number of classes;
    $|D| = N$;#the total number of samples
Output:
    A balanced training set $D'$;
$I_{resample} = int(N/C)$
for $i \leftarrow 1$ to $C$ do
    if $|D_i| < I_{resample}$ then
        $D_i' = SMOTE(D_i, I_{resample})$ #Generating new samples to minority class
    end if
    if $|D_i| > I_{resample}$ then
        $D_i' = NearMiss(D_i, I_{resample})$ #Removing values from majority class
    end if
end for
return D'
```

Algorithm 1. Two-stage resampling method of ROGONG-IDS

3.1.3. Classification decision: Extreme Gradient Boosting (XGBoost)

The XGBoost algorithm is a supervised learning method in machine learning and aims to turn weak learners into strong learners with ensemble learning (Chen & Guestrin, 2016). It is an improved version of the gradient boosting method for DTs. This algorithm aims to provide scalability in tree boosting systems, ensure efficient use of computational resources, and improve the model's performance in classification-regression problems. In the implementation phase, the initial leaf is created, then new trees are created based on the prediction errors. This continues until the number of decision trees that can be provided as hyperparameters or until the development in the model stops. Its difference from other gradient boosting methods tested for the classifier model within the scope of this study, such as GBM or LightGBM, is that it is suitable for parallel processing and is tolerant of datasets with missing data. Since it consists of many hyperparameters, the optimization phase is essential in its application for ideal hyperparameter values.

3.2. Dataset

The UNSW-NB15 dataset was used in this study. The Intelligent Security Group collected this dataset at the Australian Centre for Cyber Security (Moustafa & Slay, 2015). The research group combined the current standard network and synthetic attack data to generate this dataset. The network flow samples were stored as vectors of 49 attributes, of which two are binary, three are categorical, 37 are numerical input attributes, and 1 class attribute. It is a comprehensive dataset representing a modern network with these features.

On the other hand, there is a high level of imbalanced class problem in the dataset. 87.35% of the dataset is regular traffic, and only 12.65% is attack traffic. In the presented study, the dataset was divided into training and testing at a ratio of 7:3. *Table III* shows the attack class distributions in the dataset in detail.

Table iii
Attack class distributions

Class	Training set size	Test set size	Total
Analysis	1,874	803	2,677
Backdoor	1,630	699	2,329
DoS	11,447	4,906	16,353
Exploits	31,167	13,358	44,525
Fuzzers	16,972	7,274	24,246
Generic	150,837	64,644	215,481
Normal	1,553,134	665,630	2,218,764
Reconnaissance	9,791	4,196	13,987
Shellcode	1,058	453	1,511
Worms	122	52	174
Total (10 classes)	1,778,032	762,015	2,540,047

4. EXPERIMENTAL ANALYSIS

We used the UNSW-NB15 dataset to measure the overall performance of the ROGONG-IDS method. *Table IV* represents the system environment parameters used in this study.

Table iv
Test environment

Parameter	Environment / Version
Operating System	macOS Monterey
CPU	1,4 GHz Quad-Core Intel Core i5
GPU	Intel Iris Plus Graphics 645
Memory	16GB

4.1. Evaluation metrics

We use ACC, DR, FAR, F_1 score, and precision indicators, which are commonly used in class imbalance systems. Samples corresponding to the attack are considered positive; other samples are considered negative. The meanings of the metrics are listed below:

ACC (Accuracy): the percentage of correctly classified samples among all samples.

DR (Detection Rate): the rate of correctly predicted positive samples.

FAR (False Alarm Rate): the proportion of negative samples incorrectly evaluated as positive.

Precision: how many samples are predicted to be positive are positivesamples.

F_1 Score: the harmonic average of precision and DR parameters.

When applying multi-class classification, each class must be calculated using a weighted average method based on the number of samples in the category to understand the detection performance of the model on unbalanced data. Equations 1 - 5 represents the formula for the metrics used (TP: true positive, TN: true negative, FP: false positive, FN: false negative).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$DR = \frac{TP}{TP + FN} \tag{2}$$

$$FAR = \frac{FP}{FP + TN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F_1\ Score = \frac{2 * DR * Precision}{DR + Precision} \quad (5)$$

TP/FP and TN/FN are the numbers of the correctly and incorrectly predicted samples, respectively.

4.2. Hyper-parameter optimization

Hyperparameters can improve the performance of the model's learning process. It may be possible to reach the best performance of the model in the shortest time by adjusting the hyperparameters. Therefore, choosing the optimization method that will improve model ACC with the least time and power cost is essential. Grid search, one of the frequently used optimization methods, is a brute-force technique (Putatunda & Rama, 2018). It uses manually created subsets to optimize hyperparameters (Schaer, Müller, & Depeursinge, 2016). Although it is a simple method, increasing the number of hyperparameters increases the computational cost exponentially. It is reliable for low-dimensional spaces (Bergstra & Bengio, 2012). On the other hand, random search aims to tune hyperparameters by selecting random points in the search space (Bergstra, Bardenet, Bengio, & Kégl, 2011). It is not suitable for models with many hyperparameters like Grid search. Therefore, these two techniques, which are frequently used, are pricey for models with many hyperparameters. In this context, studies show that Hyperout outperforms Random search and Grid search in terms of ACC and time in optimizing the hyperparameters of the Extreme gradient boosting model and different machine learning models (Bergstra, Komer, Eliasmith, Yamins, & Cox, 2015; Putatunda & Rama, 2018). ROGONG-IDS uses Distributed Asynchronous Hyperparameter Optimization (Hyperopt) for hyperparameter tuning. Hyperopt, identified as a black box optimization technique (Klein, Falkner, Bartels, Hennig, & Hutter, 2017), was developed to automate hyperparameter optimization based on Bayesian optimization. Hyperopt uses Bayesian optimization to define and narrow the search space and maximize the probability function. ROGONG-IDS model ACC increased from 96.49% to 97.30% after using Hyperopt within a reasonable time. *Table V* shows the default XGBoost hyperparameters and the final hyperparameter values used after optimization.

Table v
XGBoost hyperparameters before and after from Bayesian optimization

Parameter	Value Before Optimization	Value After Optimization
Learning Rate	0.3	0.5
Number of Estimators	100	5,000
Max Depth	6	36
Colsample Bytree	1	0.61
Min Child Weight	1	4
Subsample	1	0,9

4.3. Multi-class classification

Table VI represents the DR metrics for each class. Although the ROGONG-IDS essentially uses the XGBoost algorithm, other gradient boost-based algorithms have also been tried in processing the method with the UNSW-NB15 dataset. According to the metrics shown in *Table VI*, ROGONG-IDS with XGBoost achieves the best overall performance in terms of DR, ACC, and F_1 score. These metrics are 97.30%, 98.16%, and 97.65%, respectively.

Table vi
Multi-class classification perfomance comparison between LightGBM, GBM, and XGBoost

Class	LightGBM	GBM	RXGBoost
Analysis	0.84	0.67	0.31
Backdoor	0.23	0.11	0.26
DoS	0.06	0.13	0.47
Exploits	0.46	0.48	0.54

Fuzzers	0.66	0.73	0.70
Generic	0.97	0.97	0.98
Normal	0.99	0.99	0.99
Reconnaissance	0.81	0.81	0.77
Shellcode	0.88	0.55	0.53
Worms	0.83	0	0.83
DR (%)	96.55	96.26	97.30
Accuracy (%)	96.55	96.26	97.30
Precision (%)	98.30	97.91	98.16
F_1 Score (%)	97.18	96.91	97.65
Train-Time (s)	15.08	4,336.71	205.27
Test-Time (s)	2.2	0.73	0.82

Table VII compares advanced IDS methods in the literature and the ROGONG-IDS method. With the ROGONG-IDS method, the DR metric has been improved for many classes. However, the DR value for the three classes remained below 50%. These classes are ”Analysis", “Backdoor", and “DoS" classes. When we look at the test times, it is seen that ROGONG-IDS shows the best performance in the literature. The test time is 8 seconds in the SGM method, while only 0.81 seconds in the ROGONG-IDS method.

Table vii

Comparison multi-class classification results with advanced methods on the UNSW-NB15 dataset

Class	M1	M2	M3	M4	M5	M6
Analysis	0.27	0.01	0	0.15	-	0.31
Backdoor	0.51	0	0.6	0.21	-	0.26
DoS	0.39	0	0.18	0.80	-	0.47
Exploits	0.45	0.57	0.86	0.71	-	0.54
Fuzzers	0.67	0.40	0.53	0.35	-	0.70
Generic	0.97	0.61	0.97	0.96	-	0.98
Normal	0.98	0.82	0.80	0.81	-	0.99
Reconnaissance	0.82	0.24	0.79	0.80	-	0.77
Shellcode	0.88	0,00	0.51	0.92	-	0.53
Worms	0.83	0,00	0.59	0.79	-	0.83
DR (%)	96.54	63.27	78.65	95.68	-	97.30
Accuracy (%)	96.54	89.13	78.65	89.08	85.89	97.30
Precision (%)	98.30	89.13	78.65	86.05	-	98.16
F_1 Score (%)	97.26	90.85	78.65	90.61	-	97.65
FAR	-	-	0.11	-	0.6	0.51
Train-Time (s)	47.22	-	-	-	-	205.27
Test-Time (s)	8.26	-	-	-	-	0.81

(M1: SGM-CNN (Zhang *et al.,* 2020), M2: Two stage – DL (Khan *et al*., 2019), M3: Hybrid Machine Learning (Chkirbene *et al.,* 2020), M4: ICVAE-DNN (Yang *et al*., 2019), M5: ADASYN and LightGBM (Liu *et al*., 2021), M6: ROGONG-IDS)

The validity of the robust structure of the ROGONG-IDS model, the performance of which was tested with UNSW-NB15, was evaluated with the NSL-KDD dataset used to assess many attack detection models in the literature (Tavallaee, 2009). The categorical features of the NSL-KDD training dataset, which includes different types of cyber attacks, were digitized with one-hot encoding, and the attack types were mapped in the data preprocessing stage. The classification results obtained using the resampling approach proposed within the scope of ROGONG-IDS confirm the robust structure of the model. Classification results are presented in *Table VIII*.

Table viii
Multi-class classification results on the NSL-KDD dataset

Class	**ROGONG-IDS**
Normal	0.95
DoS	0.95
Probe	0.99
R2L	0.39
U2R	0.10
DR (%)	94.31
Accuracy (%)	94.31
Precision (%)	96.67
F_1 Score (%)	95.23
FAR	0.0002
Train-Time (s)	48.41
Test-Time (s)	0.19

5. DISCUSSION

Experimental results show that the ROGONG-IDS method significantly improves the DR metric. The source of this improvement is using a two-method imbalance data module with XGBoost in the ROGONG-IDS method. According to the experimental analysis results, the XGBoost algorithm produces more successful results than other methods (GBM, LightGBM). XGBoost provided a higher DR value than the other two classifiers in attack types “Backdoor”, “DoS”, “Exploits”, “Generic”, “Normal”, and “Worms”. When examined in general, it produces more successful results than the other two algorithms based on DR, ACC, F_1 score, and test time metrics. When comparing the ROGONG-IDS method with other state-of-the-art methods, it is seen that ROGONG-IDS is the most successful IDS model in the literature in terms of DR, ACC, F_1 score, and test time. Therefore, it outperforms state-of-the-art intrusion detection methods.

The ROGONG-IDS method has been tested on two different data sets in the literature and has produced successful results in both data sets. Therefore, the method is considered to be robust.

The FAR value of the ROGONG-IDS method was observed to be lower when compared to the FAR values of similar techniques in the literature. However, ROGONG-IDS has a higher classification accuracy than these methods, which makes the approach valuable. In future studies, it is planned to improve the FAR value by keeping the classification accuracy high.

6. CONCLUSION

There are many problems in IDS that are difficult to solve. One of these problems is the datasets used in the evaluation phase. In the evaluation phase of the presented method, the UNSW-NB15 dataset, which includes the most up-to-date attack types and offers many different network parameters, was used to develop a method suitable for modern network environments. However, due to the dynamic nature of the field, it is crucial to keep the datasets up-to-date. Another problem is the case of imbalanced class. In network intrusion systems datasets, attack data items are less frequent than normal data items. This leads to an imbalance between the classes in the dataset, known as the imbalanced class problem in the literature. Increasing the data size to overcome this problem also causes an increase in the computing power and time required for data processing.

According to the evaluation findings of the proposed model, the XGBoost algorithm is more successful than other methods (GBM, LightGBM). The ROGONG-IDS model was compared with five advanced IDS models in the literature during the evaluation phase. The model’s DR, ACC, and F_1 score metrics were obtained as 97.30%, 97.30%, and 97.65%, respectively. These results prove that the ROGONG-IDS model outperforms the state-of-the-art methods. On the other hand, the ROGONG-IDS model has a fast testing time (0.81s). As a result, ROGONG-IDS is an efficient solution for real-time intrusion detection applications, delivering high success quickly. The ROGONG-IDS model could therefore be applied to areas where streaming data is imbalanced.

Appendix A. Source Code: The source codes written in Python are provided in datastd-dev/Github (2021).

Peer-review: Externally peer-reviewed.
Author Contributions: Conception/Design of Study- A.O.A., G.Ç.Ç.; Data Acquisition- A.O.A.; Data Analysis/Interpretation- A.O.A.; Drafting Manuscript- A.O.A.; Critical Revision of Manuscript- A.O.A., G.Ç.Ç.; Final Approval and Accountability- A.O.A., G.Ç.Ç.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

REFERENCES

Andresini, G., Appice, A., Mauro, N. D., Loglisci, C., & Malerba, D. (2020). Multi-Channel Deep Feature Learning for Intrusion Detection. *IEEE Access*, *8*, 53346–53359. https://doi.org/10.1109/ACCESS.2020.2980937

Belouch, M., El Hadaj, S., & Idhammad, M. (2018). Performance evaluation of intrusion detection based on machine learning using Apache Spark. *Procedia Computer Science*, *127*, 1–6. https://doi.org/10.1016/j.procs.2018.01.091

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *Advances in Neural Information Processing Systems*, *24*. Curran Associates, Inc. Retrieved from https://papers.nips.cc/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

Bergstra, J., & Bengio, Y. (2012). *Random Search for Hyper-Parameter Optimization*.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., & Cox, D. D. (2015). Hyperopt: A Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, *8*(1), 014008. https://doi.org/10.1088/1749-4699/8/1/014008

Bhavani, T. T., Rao, M. K., & Reddy, A. M. (2020). *Network Intrusion Detection System Using Random Forest and Decision Tree Machine Learning Techniques*. *1045*, 637–643. https://doi.org/10.1007/978-981-15-0029-9_50

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chkirbene, Z., Eltanbouly, S., Bashendy, M., AlNaimi, N., & Erbad, A. (2020). Hybrid Machine Learning for Network Anomaly Intrusion Detection. *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 163–170. https://doi.org/10.1109/ICIoT48696.2020.9089575

datastd-dev/Github. (2021). ROGONG-IDS/GitHub. Retrieved December 21, 2022, from https://github.com/datastd-dev/ROGONG-IDS (Original work published December 15, 2021)

Demidova, L., & Klyueva, I. (2017). SVM classification: Optimization with the SMOTE algorithm for the class imbalance problem. *2017 6th Mediterranean Conference on Embedded Computing (MECO)*, 1–4. https://doi.org/10.1109/MECO.2017.7977136

Ericsson. (2021). *Ericsson Mobility Report November 2021.*

Haibo He & Yunqian Ma. (2013). Imbalanced Learning: Foundations, Algorithms, and Applications | Wiley. Retrieved December 21, 2022, from Wiley. comwebsite: https://www.wiley.com/en-us/Imbalanced+Learning%3A+Foundations%2C+Algorithms%2C+and+Applications-p-9781118074626

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Multi-Stage Optimized Machine Learning Framework for Network Intrusion Detection. *IEEE Transactions on Network and Service Management*, *18*(2), 1803–1816. https://doi.org/10.1109/TNSM.2020.3014929

Kaja, N., Shaout, A., & Ma, D. (2019). An intelligent intrusion detection system. *Applied Intelligence*, *49*, 3235–3247. https://doi.org/10.1007/s10489-019 01436-1

Khan, F. A., Gumaei, A., Derhab, A., & Hussain, A. (2019). A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection. *IEEE Access*, *7*, 30373–30385. https://doi.org/10.1109/ACCESS.2019.2899721

Klein, A., Falkner, S., Bartels, S., Hennig, P., & Hutter, F. (2017, March 7). *Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets*. arXiv. https://doi.org/10.48550/arXiv.1605.07079

Liu, J., Gao, Y., & Hu, F. (2021). A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. *Computers & Security*, *106*, 102289. https://doi.org/10.1016/j.cose.2021.102289

Mauldin, A. (2021). *Global Internet Traffic and Capacity Return to Regularly Scheduled Programming*. Retrieved from https://blog.telegeography.com/internet-traffic-and-capacity-return-to-their-regularly-scheduled-programming

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*, 1–6. https://doi.org/10.1109/MilCIS.2015.7348942

Mulyanto, M., Faisal, M., Prakosa, S. W., & Leu, J.-S. (2021). Effectiveness of Focal Loss for Minority Classification in Network Intrusion Detection Systems. *Symmetry*, *13*(1), 4. https://doi.org/10.3390/sym13010004

Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., & Han, K. (2018). Enhanced Network Anomaly Detection Based on Deep Neural Networks. *IEEE Access*, *6*, 48231–48246. https://doi.org/10.1109/ACCESS.2018.2863036

Putatunda, S., & Rama, K. (2018). A Comparative Analysis of Hyperopt as Against Other Approaches for Hyper-Parameter Optimization of XGBoost. *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, 6–10. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3297067.3297080

Schaer, R., Müller, H., & Depeursinge, A. (2016). Optimized Distributed Hyperparameter Search and Simulation for Lung Texture Classification in CT

Using Hadoop. *Journal of Imaging*, *2*(2), 19. https://doi.org/10.3390/jimaging2020019

Sumaiya Thaseen, I., Saira Banu, J., Lavanya, K., Rukunuddin Ghalib, M., & Abhishek, K. (2021). An integrated intrusion detection system usin correlation-based attribute selection and artificial neural network. *Transactions on Emerging Telecommunications Technologies*, *32*(2), e4014. https://doi.org/10.1002/ett.4014

Tavallaee, M., Bagheri, E., Lu, W.,& Ghorbani, A. A.(2009). A detailed analysis of the KDD CUP 99 data set. *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, 2009, pp. 1-6, doi: 10.1109/CISDA.2009.5356528

Yang, Y., Zheng, K., Wu, C., & Yang, Y. (2019). Improving the Classification Effectiveness of Intrusion Detection by Using Improved Conditional Variational AutoEncoder and Deep Neural Network. *Sensors*, *19*(11), 2528. https://doi.org/10.3390/s19112528

Yen, S.-J., & Lee, Y.-S. (2006). Under-Sampling Approaches for Improving Prediction of the Minority Class in an Imbalanced Dataset. In D.-S. Huang, K. Li, & G. W. Irwin (Eds.), *Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006* (pp. 731–740). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-37256-1_89

Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. *IEEE Access*, *5*, 21954 21961. https://doi.org/10.1109/ACCESS.2017.2762418

Zhang, H., Huang, L., Wu, C. Q., & Li, Z. (2020). An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset. *Computer Networks*, *177*, 107315. https://doi.org/10.1016/j.comnet.2020.107315

Zhang, H., Wu, C. Q., Gao, S., Wang, Z., Xu, Y., & Liu, Y. (2018). An Effective Deep Learning Based Scheme for Network Intrusion Detection. *2018 24th International Conference on Pattern Recognition (ICPR)*, 682–687. https://doi.org/10.1109/ICPR.2018.8546162