DOI 10.15622/ia.23.6.10

C. NARAYANARAO, V. MANDAPATI, B. BODDU

SYNERGISTIC APPROACHES TO ENHANCE IOT INTRUSION DETECTION: BALANCING FEATURES THROUGH COMBINED LEARNING

Narayanarao C., Mandapati V., Boddu B. **Synergistic Approaches to Enhance IoT Intrusion Detection: Balancing Features through Combined Learning.**

Abstract. The Internet of Things (IoT) plays a crucial role in ensuring security by preventing unauthorized access, malware infections, and malicious activities. IoT monitors network traffic as well as device behaviour to identify potential threats and take appropriate mitigation measures. However, there is a need for an IoT Intrusion Detection system with enhanced generalization capabilities, leveraging deep learning and advanced anomaly detection techniques. This study presents an innovative approach to IoT IDS that combines SMOTE-Tomek link and BTLBO, CNN with XGB classifier which aims to address data imbalances, improve model performance, reduce misclassifications, and improve overall dataset quality. The proposed IoT IDS system, using the IoT-23 dataset, achieves 99.90% accuracy and a low error rate, all while requiring significantly less execution time. This work represents a significant step forward in IoT security, offering a robust and efficient IDS solution tailored to the changing challenges of the interconnected world.

Keywords: min-max normalization, SMOTE-Tomek Link, BTLBO algorithm, CNN with XGB, Adam Optimizer.

1. Introduction. IoT is a system that connects items that may transfer information without the necessity for interactions between humans and computers, such as computing equipment, automated and digital technologies, objects, animals, and humans. IoT essentially connects the real and virtual worlds. IoT's primary idea is to create a secure, independent link that allows info interchange between actual physical devices and applications [1]. The IoT Analytics study reveals that over 11 billion IoT devices are connected and utilized. Additionally, it is demonstrated the number of devices has increased by more than 10%. It is predicted that over 21 billion linked IoT devices will be worldwide by 2025. Due to general use in several areas and businesses, such as agriculture, transportation, logistics, and healthcare are all examples of smart cities and smart homes, the IoT has seen tremendous growth in recent years [2]. Organisations that use IoT devices in information technology systems have introduced new cybersecurity risks. These new threats call into question fundamental assumptions such as operating ecosystem security, mobility, efficiency, and safety. New danger vectors risk financial and bodily well-being and affect lives' technical components [3, 4].

The ecosystem in which IoT devices are deployed is vulnerable to several threats [5] from outsiders, including hackers, malicious software, and viruses [6]. These hackers' primary objective is to launch assaults compromising network data integrity [7]. Additionally, the intrusion may result in a denial of service (DoS) attack [8] that depletes energy in an IoT environment as well as network and device resources [9]. The author infers from the literature that many studies on the IoT employ security methods based on cryptography, such as symmetric and public key cryptosystems [10, 11].

Because IoT devices have limited resources, implementing cryptographic algorithms in IoT security leads to effective communication and processing overhead [12]. This problem can be resolved by designing and implementing intrusion detection systems (IDS). To effectively secure IoT communication, IDS has been accepted in IoT environments to guide and detect imposters [13, 14]. The detection of critical and particular threats for traditional networks and a portion of the Internet of Things networks has been taught using various Machine learning (ML) and Deep learning (DL) models for IDS [15]. IDSs currently apply similar attribute ideas to IoT devices. However, IoT devices vary in a variety of bearings, including physical characteristics, utility, potential computing power, and variable capacity aimed at generating decided appearances [16, 17]. When hubs are merged then generate data, the features become sparse as unimportant qualities are set to null values or zeros. One limitation striking the precision's effectiveness is data sparsity. A selection of features, an essential component of an ML method, contributes much to the training phase speed and finding accuracy. To enhance the identification of variations of anomalous behavior, many feature selection strategies have been developed. However, the accuracy of anomaly-based ID detection is considered a significant issue due to the constantly evolving nature of the IoT ecosystem. To achieve robust performance across the varied IoT environment, this study provides a unique technique for deep learning IDS.

The primary contribution of the proposed project is given below.

– IoT faces significant security challenges due to remote access and unreliable networks. To prevent attacks, IoT environments employ effective security management techniques and ID systems. Still, there are also possibilities to improve both accuracy and performance. The proposed novel deep learning technique in the IDS aims to address this issue.

– For the preprocessing stage, the novel approach utilizes the process of removal of null values, one hot encoding technique incorporating the numerical representation which enables the creation of a digital feature vector, Minmax normalization for dimensional removal, and Synthetic Minority Over-sampling Technique (SMOTE) Tomek technique for balancing synthetic data.

– This study employs a novel technique called Binary Teaching Learning Based Optimization (BTLBO) algorithm for feature selection.

– For feature extraction and classification, the approach uses a network as Convolution Neural Network (CNN) with extreme Gradient Boosting (XGB) classifier to predict the classes. Thus, the hybrid method accurately detects the intrusion in the IoT.

The description of the sections indicates that the article will cover existing research on IoT intrusions, the proposed methodology and results of the proposed work, and also draw conclusions and suggest some possible further research directions.

2. Literature survey. This portion of the article delves into CNN-based intrusion detection systems published in the literature.

To resolve network assault binary and multiclass categorization, in paper [18] the authors created two models based on DL and employed a CNN architecture. A hybrid two-step pre-processing method is also suggested to produce useful features. Deep feature synthesis is used in the proposed strategy to combine dimensionality reduction with feature engineering. It was shown that the multiclass models' accuracy of classification is lower than binary class models. Instead of traditional anomalous attack behaviors, the authors in paper [19] used statistical behaviors since they are simpler to reckon and extract without sacrificing performance. Because it primarily considers statistical characteristics of network traffic, the model’s accuracy for multiclass categorization is lower.

The Temporal CNN (TCNN) was proposed by the authors in paper [20] and is accumulated with SMOTE with nominal continuity to handle imbalanced datasets. To find network system abnormality, the authors in [21] used the novel CNNs binary and multiclass classification model. Even though CNN has various characteristics that make it especially suited for IDS, such as high attainment precision, finding rate, model training time, and feature selection procedures, the efficiency of ML models is improved. Hybrid models have been quite common in recent years for categorizing attacks on IoT networks. To address the IDS issues related to time consumption and inefficiency, the authors in study [22] introduced a cascade ID method that depends on distributed k-means and Random Forest.

Along with the Ant Lion optimization approach, which combines CNN and Long Short Term Memory (LSTM), study [23] introduced a new customized recurrent neural network model that is optimized for detecting intrusion. The Lion Swarm Optimization method is employed to optimize CNN hyperparameters for perfect composition for learning structural data. The authors in [24] suggested a highly accurate IDS model for valuable uprooting and learning of contiguous secular features using optimized CNN and Hierarchical multiscale LSTM. Through careful feature selection, this approach can increase detection accuracy. A Deep Capsule Network (DCN) ID model that depends on the system of attention was suggested by the authors in [25]. To increase feature extraction, the model integrates DCN, and an Attention Mechanism is employed to minder the model's attention toward qualities with substantial consequences. Two solutions are utilized to balance the dynamic powerful routing procedure after the double routing algorithm captures the characteristics in multiple directions. Because the dynamic routing contrivance of the CN consumes more time than a normal NN, the operational efficiency of CN must be increased.

To protect the computer, network nodes and data, in study [26] the authors introduced a unique Network Intrusion Detection System architecture that depends on a deep capsule neural network that creates usage of network spectrogram pictures produced utilizing the short-time Fourier transform. In comparison to previous published works, the computational complexity is higher. To identify intrusions in the IoT environment, in paper [27] the authors found a novel multi-objective evolutionary CNN for IDS. In the context of IoT and cloud computing, a new approach to IDSs was proposed in [28]. The major goal is to develop effective feature extraction and selection strategies by utilizing the widespread use of deep learning and metaheuristic optimization algorithms. An approach based on PCA and CNN was put out by the authors in paper [29] to identify intrusion in EDGE Computing. Feature selection and data balancing are not employed to improve categorization accuracy. For machine learning-based IDS, the authors in [30] gave a feature selection technique for extracting useful subsets of features based on the idea of the math concept of sets. The designed ML-based IDS system contains three stages: data pre-processing, proportions lessening and size selection, model training, and categorization.

An analysis of various works reveals that many of them neglect the multiclass imbalance distribution and feature selection techniques for improved accuracy. Unique techniques are needed to manage imbalance distribution and select the best feature set for multiclass classification.

3. Proposed methodology. IoT IDS face limitations due to data imbalances, model generalization, and performance optimization. Traditional approaches struggle to address these issues, leading to suboptimal performance and limited scalability. Imbalanced datasets can bias model training and result in poor classification performance, especially for minority intrusion classes. Current techniques for handling imbalanced data may not capture underlying patterns or introduce biases. Conventional

IDS methods also lack the ability to generalize across diverse IoT environments and adapt to evolving threats, relying on handcrafted features or simplistic anomaly detection algorithms. To overcome these limitations, our research proposes an efficient ID using a deep learning-based categorization strategy to increase the IDS's accuracy. The proposed method starts with pre-processing to eliminate duplicate instances and missing values, followed by numerical processing to produce a digital feature vector. The Min-Max Normalization approach is used for linear and uniform mapping of feature ranges, characteristics can be adjusted for faster removal of dimensions and arithmetic processing.

The problem with ML-based IDSs is using an unbalanced dataset to train a model. SMOTE-Tomek links, which combine SMOTE for artificial information for the minority class, along with Tomek Connections for excluding data identified as from the majority, solve this issue. The proposed study uses CNN with XGB for classification, which includes two convolutional hidden layers, batch normalization, Exponential Linear Unit (ELU), max-pooling, dropout layer, and Adam optimizer weights. Our work aims to enhance the performances of IoT IDs. Figure 1 depicts the overall design of the proposed technique.

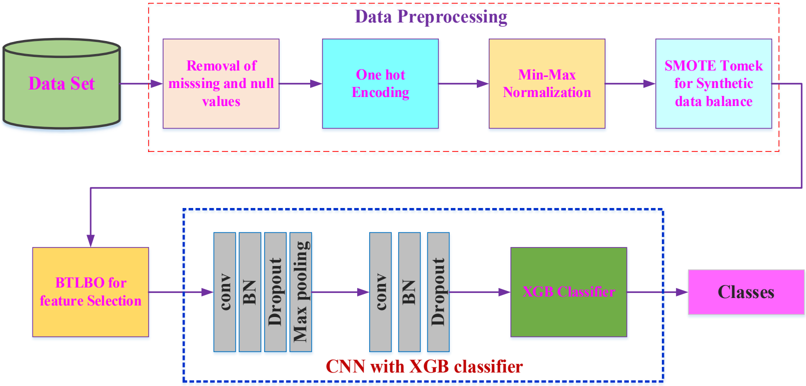


Fig. 1. Proposed Block Diagram

3.1. Data Preprocessing. It is a crucial stage in data analysis and ML work, involving cleaning, conversion, and converting raw information into suitable formats for analysis or model training. Data preparation can have a substantial impact on the accuracy and effectiveness of an analysis or model. It involves removing missing and null values from the dataset to ensure the effectiveness of the ML model, as this removes incomplete or unreliable data.

Informal information, including categorical and symbolic aspects, is handled using a one-hot encoding technique. This process converts non-numeric data into a digital feature vector, which can interpret and utilize these features effectively.

The Min-Max Normalization method is used for smooth and consistent mapping of each feature's parameter variation over a specific interval. This normalization technique improves the stability and convergence of machine learning algorithms by ensuring consistent scaling across features. It is a way that provides a balance of assessments among information obtained from prior and subsequent procedures. Features are further normalized to a Gaussian distribution, which aids in the removal of dimensions and accelerates arithmetic processing.

$$X_{new} = \frac{X - min(X)}{max(X) - min(X)}, \tag{1}$$

where $X_{new} = the\ new\ value\ from\ the\ normalized\ results$, X = old value, $Max(X) = Maximum\ value\ in\ the\ dataset$, $Min(X) = Minimum\ value\ in\ the\ dataset$.

3.1.1. Proposed SMOTE-Tomek Link. The research on strengthening IoT IDS through synergistic techniques is focused on overcoming the issues given by data imbalances and improving the overall performance of detection models. One of the innovations in this study is the use of an upgraded SMOTE-Tomek link methodology that outperforms existing methods. The SMOTE technique, when paired with Tomek links, helps to balance the dataset by producing synthetic samples for the minority classes and deleting overlapping samples between the classes. This method is unique because it balances the data, guaranteeing that the created synthetic samples contribute positively to the model's learning process.

This technique is a new approach to ML-based IDSs, which addresses the issue of unbalanced datasets, ensuring accurate identification of the minority class instances which combines the SMOTE to generate synthetic data for the minority class and Tomek Links [46] to remove Tomek links from the majority class, resulting in more accurate model training. This strategic combination of preprocessing techniques contributes to the overall efficacy and reliability of the IDS model, making it more suitable for real-world scenarios with class imbalances. The use of this technique is critical to overcoming the challenge of an imbalanced dataset. The proposed technique addresses class imbalance by combining SMOTE for synthetic data generation and Tomek Links for strategic data removal. This results in more robust and accurate IDS model training.

SMOTE-Tomek combines these two techniques in the following way.

1. Start with the original dataset, which may be imbalanced.
2. Apply SMOTE to oversample the minority class, creating synthetic instances for it.
3. After SMOTE, the dataset may still contain the Tomek link, it denotes the proximity of two instances from different classes.
4. Locate and eliminate any Tomek linkages from a dataset.

The SMOTE-Tomek link aims to enhance dataset balance by oversampling the minority classes and removing noisy samples, enhancing classification model performance. It is particularly effective in imbalanced datasets and can improve machine learning models. However, it may not be suitable for all classification problems. The used resampling technique is determined by the dataset's properties and the machine learning task objectives. The effectiveness of SMOTE-Tomek should be evaluated through cross-validation and relevant performance metrics. After balancing the data with class-wise sampling, it should be proceeded to the process of feature selection by using BTLBO.

3.2. BTLBO algorithm. It is a powerful technique for feature selection in machine learning and data analysis. It optimizes the subset of relevant features used for training models, focusing on refining model performance and interpretability. Drawing inspiration from the teaching-learning process, BTLBO uses a binary encoding scheme to efficiently explore the solution space, evaluating, and selecting features that significantly contribute to the model's predictive power. BTLBO is a novel and efficient approach to feature selection, overcoming limitations in traditional methods such as high-dimensional datasets, computational inefficiency, and suboptimal search strategies. Its binary representation and amalgamation of teaching and learning strategies enable systematic evaluation of feature subsets, enhancing model accuracy and interpretability in ML applications. This innovative method offers a robust solution for addressing the challenges faced by traditional methods.

Let M_i be the mean, and T_i be the teacher at any iteration i. T_i will try to move mean M_i towards its own level. First students are trained with the assistance of a teacher. Assume that ‘s’ represents the number of features in each iteration k. Attributes, {f=1,2,….s}, ‘t’ is the number of instances i.e., population, individuals, {i=1,2,…t}.

BTLBO is a teaching-learning optimization algorithm that stimulates the teaching and learning process in a classroom. It involves all students as a population, with subjects similar to decision variables. The best learner is considered a teacher, who transfers knowledge to all learners. The learner's performance is based on the fitness value of the individual in the population. TLBO operates in two phases: teacher and learner (Figure 2).

Algorithm 1. BTLBO [42]

Step 1: Initialize several instances (binary population), several subscripts and $X_{f,i,k}$ and an end condition.

Step 2: Calculate the mean for learners as $M_{f,k}$

Step 3: Using equation (2), determine the fitness of people.

$$\text{Fitness } (X_{f,i,k}) = \text{Accuracy } (X_{f,i,k}). \tag{2}$$

(Teacher Phase)

Step 4: Upgrade students with the assistance of the instructor. i.e. teacher phase

(a) Choose the highest fitness value from the group as a teacher.

(b) Calculate the mean variation for all traits concerning the best individual as shown in the equation

$$\text{Diff\_}Mean_{f,i,k} = r_k(X_{f,i,best,k} - T_f M_{f,k}), \tag{3}$$

where $X_{f,i,best,k}$ the best individual in f. T_f, teaching factor with the value 1 or 2, r_k is the random number ranging from 0 to 1.

(c) The best person serves as a teacher and mentor to others.

(d) $X'_{f,i,k} = 0, if\ X_{f,i,k} + \text{ Diff\_}Mean_{f,k} < 0.5$

$$X'_{f,i,k} = 1, if\ X_{f,i,k} + \text{ Diff\_}Mean_{f,k} \geq 0.5, \tag{4}$$

where $X'_{f,i,k}$ the trained value of $X_{f,i,k}$

If the result $X'_{f,i,k}$ is better than $X_{f,i,k}$, Otherwise, replace the previous value with the new value.

Step5: updates each learner with the assistance of other learners using the eq (5,6)

(Learner Phase)

(a) Select two cases U and V that satisfy the criterion $X'_{total-U,k} \neq X'_{total-V,k}$ at random, Where X_(total-U,k)^' , X_(total-V,k)^' of U and V respectively

(b) If $X'_{total-U,k}$ is better than $X'_{total-V,k}$

$$X''_{f,U,k} = 1 \text{ if } X'_{f,U,k} + r_k(X'_{f,U,k} - X'_{f,V,k}) \geq 0.5). \tag{5}$$

Or

$$\begin{aligned} X''_{f,U,k} &= 0 \text{ if } X'_{f,U,k} + r_k + (X'_{f,V,k} - X'_{f,U,k}) < 0.5. \\ X''_{f,U,k} &= 1 \text{ if } X'_{f,U,k} + r_k + (X'_{f,V,k} - X'_{f,U,k}) \geq 0.5. \end{aligned} \tag{6}$$

(c) If X_(f,U,k)^"= is better than X_(f,U,k)^', then continue the prior value, otherwise, substitute the previous value.

Step 6: if the stop condition is pleased, then report the result, then go the second step

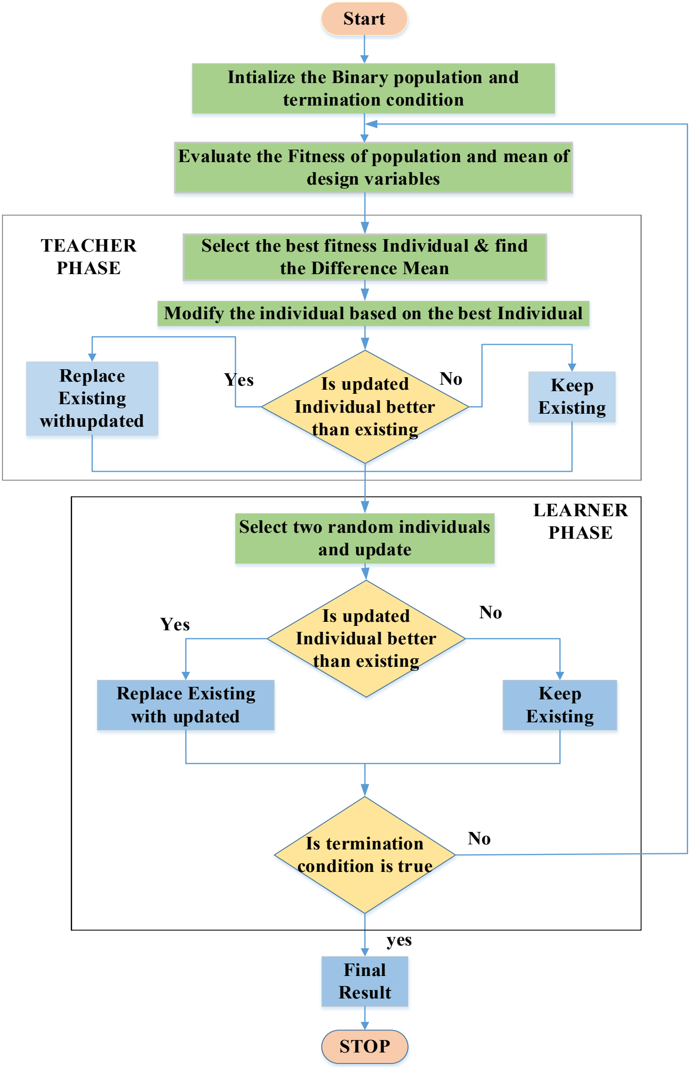

Start
Intialize the Binary population and termination condition
Evaluate the Fitness of population and mean of design variables
TEACHER PHASE
Select the best fitness Individual & find the Difference Mean
Modify the individual based on the best Individual
Replace Existing withupdated
Yes
Is updated Individual better than existing
No
Keep Existing
LEARNER PHASE
Select two random individuals and update
Yes
Is updated Individual better than existing
No
Replace Existing with updated
Keep Existing
Is termination condition is true
No
yes
Final Result
STOP

Fig. 2. BTLBO Algorithm Flow Chart [40]

This approach uses binary bits 1 and 0 to signify the presence or absence of a characteristic in a population, with the length of the binary string corresponding to the number of levels in each dataset. In the instructor phase, the mean value reflects the likelihood of a feature's appearance, while the difference mean describes learners' variation. The teacher with the highest accuracy is named. Wrapper-based feature selection approaches use predictive models to evaluate population fitness, with classification accuracy as a fitness value. The teaching factor will be chosen at random from 1 to 2. The classification accuracy can be expressed as

$$\text{CA}= \text{correct classified Instances/Total Instances.} \tag{2}$$

An individual with the lowest error rate or the highest accuracy will determine the last solution of optimal characteristics selection. The individuals added to the dataset are utilized to teach NN to boost effectiveness. The process for the optimum number of iterations is executed, and the method flow as shown in Figure 2 is presented. The BTLBO algorithm was created in the study to select an optimal subset of characteristics from an extensive database. Once features are selected, they are then refined into a categorization pipeline using CNN and the XGB classifier, combining the strengths of both techniques to improve classification performance.

3.3. Hybrid CNN-XGB Classifier. The proposed method combines CNN with the XGB classifier, resulting in a powerful combination for effective classification tasks. This hybrid approach is chosen to take advantage of CNN's capabilities in capturing intricate spatial hierarchies and patterns from complex data, such as images or sequences. CNN excels at automatic feature extraction, producing high-level representations required for reliable classification [44]. The subsequent use of the XGB classifier adds a layer of ensemble learning, allowing for efficient handling of nonlinear relationships and improving overall model performance. The proposed method enhances classification accuracy and interpretability by using CNN for feature extraction and XGB for ensemble-based classification. This approach addresses limitations in previous research, such as high-dimensional data handling and complex feature relationships, by combining deep learning and gradient boosting for optimal predictive accuracy. The CNN structure is in Figure 3.

The input layer is followed by another convolutional and pooling layer, which is a sub-sampling layer. The pooling layer then pools relevant features and performs the extraction function, and the unused features are clarified out in convolutional and pooling layers. This work presents a new intelligent deep classification algorithm using the CNN algorithm with IF…THEN rules. [45] The CNN XG classifier conducts fusion and maximal pooling operations, representing convolution for a pair of functions f, g using an integral equation for the operator t.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau. \quad (3)$$

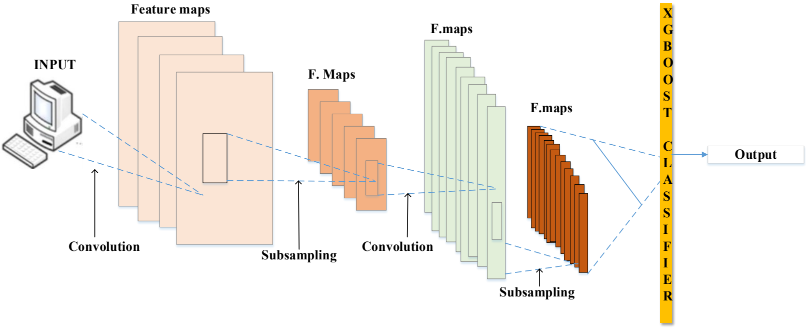


Fig. 3. Representation of the combined proposed technique

The CNN consists of dual convolutions, two batch normalizations, two dropouts and one max pooling layer, all of which are designed to automatically capture hierarchical representations from input data. Once the CNN has extracted these complex features, the feature set is fed into the XGB classifier for the final classification stage.

$$f(x) = \left[\frac{(x+1)}{x}\right]. \quad (4)$$

The XGB classifier excels at leveraging ensemble learning via gradient boosting, which creates a series of decision trees that collectively improve the model's predictive capabilities. The features extracted by the CNN serve as high-level representations of IoT data, capturing patterns and complexities that are critical for distinguishing between normal and anomalous network behaviour. The XGB classifier, which is adept at handling complex relationships and non-linearities, refines these features using a boosting process. The XGB algorithm uses CNN-extracted features to refine decision boundaries in a classification workflow. Each decision tree in the ensemble contributes to the overall classification decision, combining the strengths of multiple weak learners. This integrated approach ensures the model learns intricate features and refines them through the ensemble-based learning strategy of XGB. The final classification output distinguishes between normal and malicious IoT activities, providing a reliable and adaptive intrusion detection system for IoT environments.

4. Results and discussion. The study highlights the effectiveness of the synergistic approach to IoT Intrusion Detection System, which integrates

SMOTE-Tomek link, BTLO, Convolutional Neural Network, and XGB classifier. This section provides the dataset description, performance of the proposed method, evaluation metrics, and comparison. During the process, a Python tool is used for implementation with the tensor flow library.

4.1. Dataset description. The IoT-23 [41], [43] dataset is a crucial resource for IoT security and intrusion detection research, derived from real-world devices, simulated environments, and network traffic captures. It provides comprehensive insights into IoT operations, including network traffic attributes, device-specific information, and normal behavior. The dataset is essential for model training and evaluation, as instances are labelled to indicate potential intrusions. The analysis is carried out using IoT-23 dataset consists of 20 malware catches accomplished in IoT devices and 3 captures for benign IoT device traffic. The IoT-23 dataset consists of twenty-three different IoT network traffic recordings called scenarios. These scenarios are divided into 20 network captures of pcap files from infected IoT devices in the name of the malware executed on each scenario and three network captures of real IoT device network traffic. The dataset, which covers a wide range of anomalies in IoT ecosystems, is crucial for robust intrusion detection models due to its potential for class imbalance, a common challenge in real-world datasets.

4.2. Experimental Settings. For the evaluation of this research, the total number of samples is (1211513, 31). After preprocessing by removing the zeros and null value, the number of samples is (981934, 31). The proposed technique split the samples into 75% (736450, 31) for training and 25% (245484, 31) for testing. The number of class instances was used to calculate class weights, so the class with the fewest instances will have a high weight. Each CNN model was trained with a batch size of 32 and 10 iterations using the Adam optimizer with a learning rate of 0.001 for 100 epochs. Early halting reduces the possibility of excessive fitting, which happens when a model is refined over an abundance of eras. The batch size increased and a number of epochs lowered to see if the model's accuracy improved. For training and validation sets, the precision and loss of each model were evaluated at each epoch value.

4.3. Metrics for evaluation. The performance metrics are evaluated using the following values:

Accuracy:

$$Accuracy = \frac{TP + TN}{number\ of\ all\ samples\ in\ the\ testing\ sets}. \tag{5}$$

Precision:

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

Recall:

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

The value recall is equal to the sensitivity value.

Specificity: relates to how successfully a classifier can identify bad outcomes.

$$Specificity = \frac{TN}{TN + FP}. \quad (8)$$

The value of specificity is equal to the True Negative Rate (TNR) value.

F 1:

$$Fmeasure = \frac{2 \times precision \times recall}{precision+recall} = \frac{2TP}{2TP+FP+FN}. \quad (9)$$

PPV**:** termed a positive predictive value, which is calculated by

$$PPV = \frac{TP}{TP + FP}. \quad (10)$$

NPV**:** negative predictive value, which is calculated by

$$NPV = \frac{TN}{TN + FN}, \quad (11)$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative.

4.4. Evaluation Findings and Comparisons. To detect intrusions, the experiment used the CNN_XGG algorithm. Sensitivity, Specificity, PPV, and NPV results for multiTable.1 provides the experimental result of the proposed work which was compared with three different learning-based IDS models that work in the IoT-23. Each subset is used to evaluate CNN_LSTM, CNN_BiLSTM, CNN_GRU, and also proposed CNN_XGB models. The accuracy, precision, recall, and F1, Sensitivity, Specificity, PPV, and NPV score, of the IoT-23 dataset using CNN_XGB by comparing with the existing ranking algorithm [37] of these models are presented in Figure 6. CNN_XGB model performs better than existing models. A single hidden layer CNN successfully classified normal and abnormal situations in the IoT-23 dataset [37], demonstrating its ability to learn meaningful patterns from network traffic data, making this result impressive in detecting normal and anomalous occurrences in the IoT-23 dataset.

In the training phase, the weights for classes were computed according to the number of occurrences for every group; the minority class with a small number of instances will receive better priority. SMOTE followed to correct the class imbalances. The evaluation of the proposed model is shown in Table 1.

Table 1. Performance assessment of the proposed model

Metrics	Proposed model
Accuracy	99.90 %
Precision	99.51 %
Recall	99.95 %
F1	99.20 %
Sensitivity	99.91 %
Specificity	100 %
PPV	99.92 %
NPV	100 %
Error rate	0.012

The novel model was validated by separating the dataset presenting accuracy performances of CNN and XGB algorithms in Figure 4, the training loss of the DL algorithm mentions a link between training loss and the number of epochs in the proposed work (Figure 5). The comparative evaluation of IoT Intrusion Detection Systems is shown in Table 2.

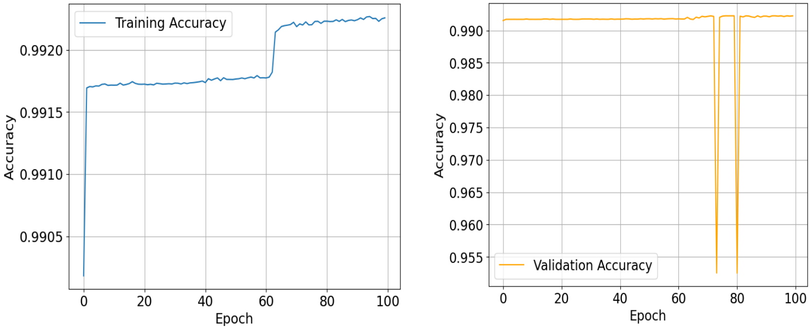


Fig. 4. Training/Validation accuracy

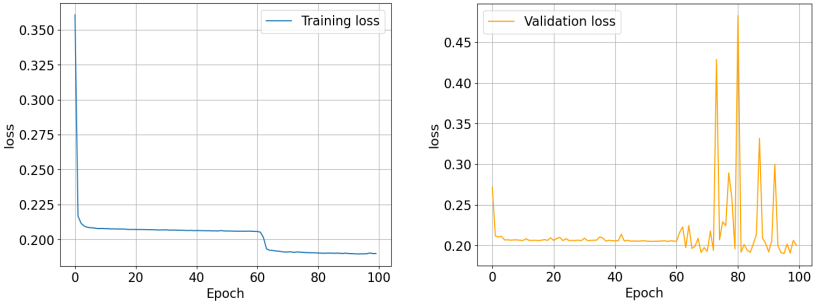
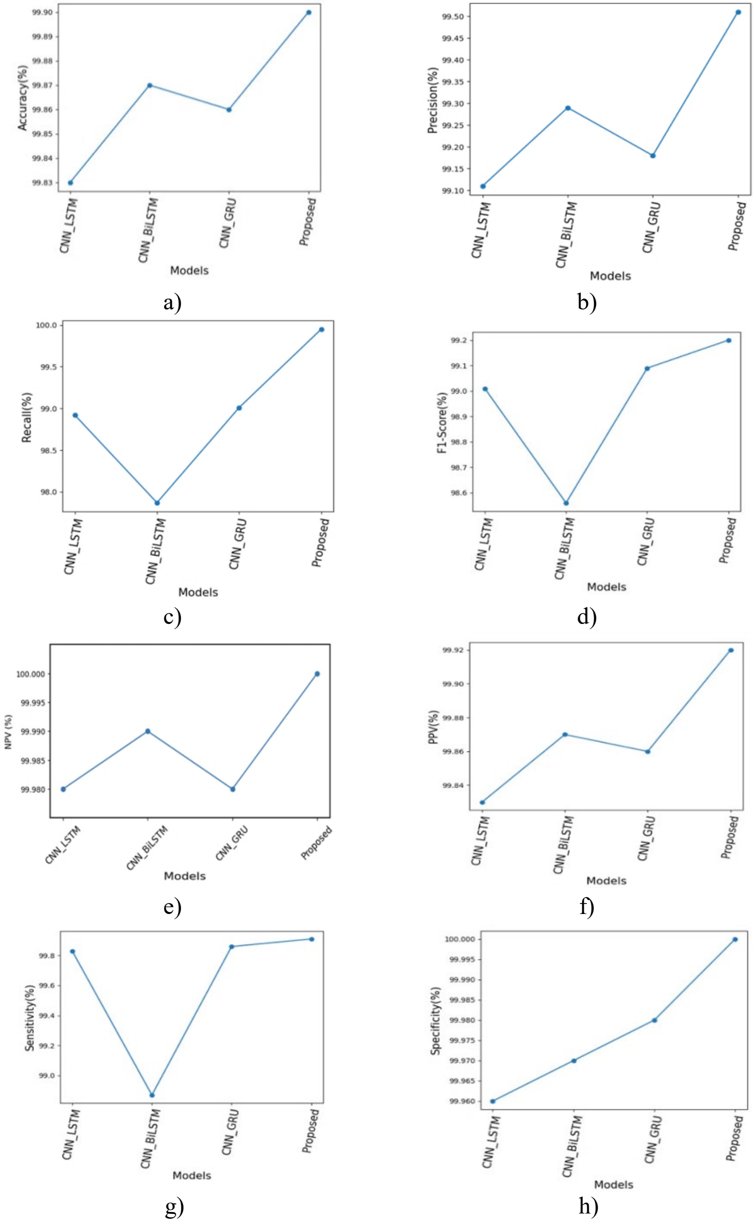


Fig. 5. Training Vs validation loss

Table 2. Performance Comparison of IoT IDS Approaches of multi-class classification

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
CNN_LSTM[37]	99.83	99.11	98.92	99.01	99.83	99.96	99.83	99.98
CNN_BiLSTM[37]	87.99	99.29	97.87	98.56	98.87	99.97	99.87	99.99
CNN_GRU[37]	86.99	99.18	99.01	99.09	99.86	99.98	99.86	99.98
Proposed	99.90	99.51	99.95	99.20	99.91	100	99.92	100

Table 2 shows that the proposed DL-based IDS models obtained better performance in identifying various forms of cyberattacks compared to existing features. In research paper [38], the RNN model gained an accuracy of 98.31%, so it was concluded that a DL model might considerably improve accuracy, permitting efficient security against threats in IoT systems. The suggested CNN model's accuracy was very low. However, the proposed model CNN_XGB showed high accuracy compared to other multi-class classifiers. From Table 3, it can be seen that the accuracy rates of DL-based IDS models are comparable, which shows the proposed model achieves the lowest error rate among IDS models [39] belonging to the CNN_XG ensemble by 0.012.



a) b) c) d) e) f) g) h)

Fig. 6. Comparison Results: a) Accuracy; b) Precision; c) Recall; d) F1-score; e) NPV; f) PPV; g) Sensitivity; h) Specificity

Table 3. Performance of the system multiclass classification and error rate variations

Algorithm	Accuracy (%)	Error Rate
CNN_LSTM[39]	99.83	0.092
CNN_BiLSTM[39]	87.99	0.1
CNN_GRU[39]	86.99	0.016
Proposed	99.90	0.012

As a result, it is proposed to balance the dataset. To address this issue, the oversampling approach was utilized to balance the datasets. Synthetic samples for the minority class are generated using SMOTE-Tomek for regional expertise instead of undefined facts regarding the faction category. The model successfully reflects the dimensional and secular connection of normal spotting challenges. The proposed methodology can be used to detect and evaluate anomalies in a wide range of IoT applications and data. Thus, CNN_XGB is capable of dealing with huge amounts of data which performs superior when dealing with huge quantities of information.

5. Conclusion. Through the creation of an IDS, this research article proposed a creative approach to improve the security of IoT environments. The observation is carried out on the IoT-23 dataset, and the results show that the proposed technique achieves good performance. The findings of the proposed work show that this combined learning technique with a balanced high-performing feature selection method, SMOTE-Tomek, CNN_XGBoost, and Adam Optimizer achieved a high accuracy of 99.90%. The integration of the proposed model contributes to constructing a strong and scalable IDS that can be applied to various IoT scenarios. Thus, our research work has practical applications and paves the way for further innovation in IoT security, ultimately contributing to the growth of secure and resilient IoT ecosystems. Future work should integrate the IDS with SIEM (Security Information and Event Management) solutions to give a more comprehensive security ecosystem for IoT networks. This can enhance the system's ability to correlate events and provide a holistic view of security.

References

1. Chopra K., Gupta K., Lambora A. Future internet: The internet of things-a literature review. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, 2019. pp. 135–139.
2. Apostol I., Preda M., Nila C., Bica I. IoT botnet anomaly detection using unsupervised deep learning. Electronics. 2021. vol. 10(16). DOI: 10.3390/electronics10161876.

3. Raghuvanshi A., Singh U.K. WITHDRAWN: Internet of Things for smart cities-security issues and challenges. 2020. DOI: 10.1016/j.matpr.2020.10.849.
4. Lokhande M.P., Patil D.D., Patil L.V., Shabaz M. Machine-to-machine communication for device identification and classification in secure telerobotics surgery. Security and communication networks. 2021. no. 1. pp. 1–16. DOI: 10.1155/2021/5287514.
5. Butun I., Osterberg P., Song H. Security of the Internet of Things: Vulnerabilities, attacks, and countermeasures. IEEE Communications Surveys and Tutorials. 2019. vol. 22(1). pp. 616–644.
6. Zahra S.R., Chishti M.A. Ransomware and internet of things: A new security nightmare. In 2019 9th international conference on cloud computing, data science & engineering (confluence). IEEE, 2019. pp. 551–555.
7. Makhdoom I., Abolhasan M., Lipman J., Liu R.P., Ni W. Anatomy of threats to the internet of things. IEEE communications surveys and tutorials. 2018. vol. 21(2). pp. 1636–1675.
8. Liang L., Zheng K., Sheng Q., Huang X. A denial of service attack method for an IoT system. In 8th international conference on Information Technology in Medicine and Education (ITME). IEEE, 2016. pp. 360–364.
9. Gray C., Ayre R., Hinton K., Tucker R.S. Power consumption of IoT access network technologies. In IEEE International Conference on Communication Workshop (ICCW). IEEE, 2015. pp. 2818–2823.
10. Gormuş S., Aydın H., Ulutaş G. Security for the internet of things: a survey of existing mechanisms, protocols and open research issues. Journal of the Faculty of Engineering and Architecture of Gazi University. 2018. vol. 33(4). pp. 1247–1272.
11. Carracedo J.M., Milliken M., Chouhan P.K., Scotney B., Lin Z., Sajjad A., Shackleton M. Cryptography for security in IoT. In Fifth International Conference on Internet of Things: Systems, Management and Security. IEEE, 2018. pp. 23–30.
12. Karati A., Fan C.I., Hsu R.H. Provably secure and generalized signcryption with public verifiability for secure data transmission between resource-constrained IoT devices. IEEE Internet of Things Journal. 2019. vol. 6(6). pp. 10431–10440.
13. Fang D., Qian Y., Hu R.Q. A flexible and efficient authentication and secure data transmission scheme for IoT applications. IEEE Internet of Things Journal. 2020. vol. 7(4). pp. 3474–3484.
14. Chaabouni N., Mosbah M., Zemmari A., Sauvignac C., Faruki P. Network intrusion detection for IoT security based on learning techniques. IEEE Communications Surveys and Tutorials. 2019. vol. 21(3). pp. 2671–2701.
15. Aldweesh A., Derhab A., Emam A.Z. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems. 2020. vol. 189(5). DOI: 10.1016/j.knosys.2019.105124.
16. Albulayhi K., Sheldon F.T. An adaptive deep-ensemble anomaly-based intrusion detection system for the internet of things. In 2021 IEEE World AI IoT Congress (AIIoT). IEEE, 2021. pp. 0187–0196.
17. Alrubayyi H., Goteng G., Jaber M., Kelly J. Challenges of malware detection in the IoT and a review of artificial immune system approaches. Journal of Sensor and Actuator Networks. 2021. vol. 10(4). DOI: 10.3390/jsan10040061.
18. Al-Turaiki I., Altwaijry N. A convolutional neural network for improved anomaly-based network intrusion detection. Big Data. 2021. vol. 9(3). pp. 233–252.
19. Lam N.T. Detecting unauthorized network intrusion based on network traffic using behavior analysis techniques. International Journal of Advanced Computer Science and Applications. 2021. vol. 12(4). DOI: 10.14569/IJACSA.2021.0120407.
20. Aljumah A. IoT-based intrusion detection system using convolution neural networks. PeerJ Computer Science. 2021. vol. 7. DOI: 10.7717/peerj-cs.721.

21. Akhtar M.S., Feng T. Deep learning-based framework for the detection of cyberattack using feature engineering. Security and Communication Networks, 2021. no. 1. DOI: 10.1155/2021/6129210.

22. Liu C., Gu Z., Wang J. A hybrid intrusion detection system based on scalable K-means+ random forest and deep learning. IEEE Access. 2021. vol. 9. pp. 75729–75740.

23. Thilagam T., Aruna R. Intrusion detection for network based cloud computing by custom RC-NN and optimization. ICT Express. 2021. vol. 7(4). pp. 512–520.

24. Kanna P.R., Santhi P. Unified deep learning approach for efficient intrusion detection system using integrated spatial-temporal features. Knowledge-Based Systems. 2021. vol. 226. DOI: 10.1016/j.knosys.2021.107132.

25. Yin S.L., Zhang X.L., Liu S. Intrusion detection for capsule networks based on dual routing mechanism. Computer Networks. 2021. vol. 197. DOI: 10.1016/j.knosys.2021.107132.

26. Khan A.S., Ahmad Z., Abdullah J., Ahmad F. A spectrogram image-based network anomaly detection system using deep convolutional neural network. IEEE access. 2021. vol. 9. pp. 87079–87093.

27. Chen Y., Lin Q., Wei W., Ji J., Wong K.C., Coello C.A.C. Intrusion detection using multi-objective evolutionary convolutional neural network for Internet of Things in Fog computing. Knowledge-Based Systems. 2022. vol. 244. DOI: 10.1016/j.knosys.2022.108505.

28. Dahou A., Abd Elaziz M., Chelloug S.A., Awadallah M.A., Al-Betar M.A., Al-Qaness M.A., Forestiero A. 2022. Intrusion detection system for IoT based on deep learning and modified reptile search algorithm. Computational Intelligence and Neuroscience. 2022. no. 1. DOI: 10.1155/2022/6473507.

29. Haq M.A., Rahim Khan M.A., AL-Harbi T. Development of PCCNN-based network intrusion detection system for EDGE computing. Computers, Materials and Continua. 2022. vol. 71(1). DOI: 10.32604/cmc.2022.018708.

30. Albulayhi K., Abu Al-Haija Q., Alsuhibany S.A., Jillepalli A.A., Ashrafuzzaman M., Sheldon F.T. IoT intrusion detection using machine learning with a novel high performing feature selection method. Applied Sciences. 2022. vol. 12(10). DOI: 10.3390/app12105015.

31. Stoyanova M., Nikoloudakis Y., Panagiotakis S., Pallis E., Markakis E.K. A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Communications Surveys and Tutorials. 2020. vol. 22(2). pp. 1191–1221.

32. Henderi H., Wahyuningsih T., Rahwanto E. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. International Journal of Informatics and Information Systems. 2021. vol. 4(1). pp. 13–20.

33. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002. vol. 16. pp. 321–357.

34. Allam M., Nandhini M. Optimal feature selection using binary teaching learning based optimization algorithm. Journal of King Saud University-Computer and Information Sciences. 2022. vol. 34(2). pp. 329–341.

35. Smys S., Basar A., Wang H. Hybrid intrusion detection system for internet of things (IoT). Journal of ISMAC 2020. vol. 2(04). pp. 190–199.

36. Raichura M., Chothani N., Patel D. Efficient CNN-XGBoost technique for classification of power transformer internal faults against various abnormal conditions. IET Generation, Transmission and Distribution. 2021. vol. 15(5). pp. 972–985.

37. Ullah I., Mahmoud Q.H. Design and development of RNN anomaly detection model for IoT networks. IEEE Access. 2022. vol. 10. pp. 62722–62750.
38. Susilo B., Sari R.F. Intrusion detection in IoT networks using deep learning algorithm. Information. 2020. vol. 11(5). DOI: 10.3390/info11050279.
39. Soliman S., Oudah W., Aljuhani A. Deep learning-based intrusion detection approach for securing industrial Internet of Things. Alexandria Engineering Journal. 2023. vol. 81. pp. 371–383.
40. Khelil H., Brahimi M. Toward an efficient web service composition based on an improved BTLBO algorithm. The Journal of Supercomputing. 2024. vol. 80(7). pp. 8592–8613.
41. Ullah I., Mahmoud Q.H. A framework for anomaly detection in IoT networks using conditional generative adversarial networks. IEEE Access. 2021. vol. 9. pp. 165907–165931.
42. Khuat T.T., Le M.H. Binary teaching–learning-based optimization algorithm with a new update mechanism for sample subset optimization in software defect prediction. Soft Computing. 2019. vol. 23(20). pp. 9919–9935.
43. Nazir A., He J., Zhu N., Qureshi S.S., Qureshi S.U., Ullah F., Wajahat A., Pathan M.S. (2024). A deep learning-based novel hybrid CNN-LSTM architecture for efficient detection of threats in the IoT ecosystem. Ain Shams Engineering Journal. 2024. vol. 15. no. 7. DOI: 10.1016/j.asej.2024.102777.
44. Gao X., Jamil N., Ramli M.I., Ariffin S.M.Z.S.Z. A Comparative Analysis of Combination of CNN-Based Models with Ensemble Learning on Imbalanced Data. JOIV: International Journal on Informatics Visualization. 2024. vol. 8. no. 1. pp. 456–464.
45. Zawaideh F.H., Al-Asad G., Swaneh G., Batainah S., Bakkar H. Intrusion Detection System for (IoI) Networks Using Convolutional Neural Network (CNN) and Xgboost Algorithm. Journal of Theoretical and Applied Information Technology. 2024. vol. 102(4). pp. 1750–1759.
46. Swana E.F., Doorsamy W., Bokoro P. Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset. Sensors. 2022. vol. 22(9). DOI: 10.3390/s22093246.

Narayanarao Chokkapu — Research scholar, Department of computer science and engineering, GITAM School of Technology. Research interests: IoT security, deep learning. The number of publications — 3. nchokkap@gitam.in; GITAM Visakhapatnam Campus, Gandhi Nagar, Rushikonda, 530045, Visakhapatnam, Andhra Pradesh, India; office phone: +91(08912)790-501.

Mandapati Venkateswara Rao — Professor, Department of computer science and engineering, GITAM School of Technology. Research interests: robotics, cloud computing. The number of publications — 10. vmandapa@gitam.edu; GITAM Visakhapatnam Campus, Gandhi Nagar, Rushikonda, 530045, Visakhapatnam, Andhra Pradesh, India; office phone: +91(08912)790-501.

Boddu Bhaskara Rao — Associate professor, Department of computer science and engineering, GITAM School of Technology. Research interests: machine learning, data science, semantic web. The number of publications — 8. bboddu@gitam.edu; GITAM Visakhapatnam Campus, Gandhi Nagar, Rushikonda, 530045, Visakhapatnam, Andhra Pradesh, India; office phone: +91(08912)790-501.

УДК 004 **DOI** 10.15622/ia.23.6.10

Ч. Нараянарао, В. Мандапати, Б. Бодду

СИНЕРГЕТИЧЕСКИЕ ПОДХОДЫ К УЛУЧШЕНИЮ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ В ИНТЕРНЕТ ВЕЩЕЙ (IOT): БАЛАНСИРОВКА ХАРАКТЕРИСТИК С ПОМОЩЬЮ КОМБИНИРОВАННОГО ОБУЧЕНИЯ

Нараянарао Ч., Мандапати В., Бодду Б. **Синергетические подходы к улучшению обнаружения вторжений в Интернет вещей (IoT): балансировка характеристик с помощью комбинированного обучения.**

Аннотация. Интернет вещей (IoT) играет важную роль в обеспечении безопасности, предотвращая несанкционированный доступ, заражения вредоносным ПО и злонамеренные действия. IoT отслеживает сетевой трафик, а также поведение устройств для выявления потенциальных угроз и принятия соответствующих мер противодействия. Тем не менее, существует потребность в системе обнаружения вторжений (IDS) IoT с улучшенными возможностями обобщения, использующей глубокое обучение и передовые методы обнаружения аномалий. В этом исследовании представлен инновационный подход к IoT IDS, который сочетает в себе SMOTE-Tomek и BTLBO, CNN с XGB классификатором, который направлен на устранение дисбаланса данных, повышение производительности модели, снижение количества неправильных классификаций и улучшение общего качества набора данных. Предложенная система обнаружения вторжений IoT, используя набор данных IoT-23, достигает 99,90% точности и низкого уровня ошибок, требуя при этом существенно меньше времени выполнения. Эта работа представляет собой значительный шаг вперед в области безопасности IoT, предлагая надежное и эффективное решение IDS, адаптированное к меняющимся проблемам взаимосвязанного мира.

Ключевые слова: минимаксная нормализация, SMOTE-Tomek Link, алгоритм BTLBO, CNN с XGB, оптимизатор Adam.

Литература

1. Chopra K., Gupta K., Lambora A. Future internet: The internet of things-a literature review. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, 2019. pp. 135–139.
2. Apostol I., Preda M., Nila C., Bica I. IoT botnet anomaly detection using unsupervised deep learning. Electronics. 2021. vol. 10(16). DOI: 10.3390/electronics10161876.
3. Raghuvanshi A., Singh U.K. WITHDRAWN: Internet of Things for smart cities-security issues and challenges. 2020. DOI: 10.1016/j.matpr.2020.10.849.
4. Lokhande M.P., Patil D.D., Patil L.V., Shabaz M. Machine-to-machine communication for device identification and classification in secure telerobotics surgery. Security and communication networks. 2021. no. 1. pp. 1–16. DOI: 10.1155/2021/5287514.
5. Butun I., Osterberg P., Song H. Security of the Internet of Things: Vulnerabilities, attacks, and countermeasures. IEEE Communications Surveys and Tutorials. 2019. vol. 22(1). pp. 616–644.
6. Zahra S.R., Chishti M.A. Ransomware and internet of things: A new security nightmare. In 2019 9th international conference on cloud computing, data science & engineering (confluence). IEEE, 2019. pp. 551–555.

7. Makhdoom I., Abolhasan M., Lipman J., Liu R.P., Ni W. Anatomy of threats to the internet of things. IEEE communications surveys and tutorials. 2018. vol. 21(2). pp. 1636–1675.
8. Liang L., Zheng K., Sheng Q., Huang X. A denial of service attack method for an IoT system. In 8th international conference on Information Technology in Medicine and Education (ITME). IEEE, 2016. pp. 360–364.
9. Gray C., Ayre R., Hinton K., Tucker R.S. Power consumption of IoT access network technologies. In IEEE International Conference on Communication Workshop (ICCW). IEEE, 2015. pp. 2818–2823.
10. Gormuş S., Aydın H., Ulutaş G. Security for the internet of things: a survey of existing mechanisms, protocols and open research issues. Journal of the Faculty of Engineering and Architecture of Gazi University. 2018. vol. 33(4). pp. 1247–1272.
11. Carracedo J.M., Milliken M., Chouhan P.K., Scotney B., Lin Z., Sajjad A., Shackleton M. Cryptography for security in IoT. In Fifth International Conference on Internet of Things: Systems, Management and Security. IEEE, 2018. pp. 23–30.
12. Karati A., Fan C.I., Hsu R.H. Provably secure and generalized signcryption with public verifiability for secure data transmission between resource-constrained IoT devices. IEEE Internet of Things Journal. 2019. vol. 6(6). pp. 10431–10440.
13. Fang D., Qian Y., Hu R.Q. A flexible and efficient authentication and secure data transmission scheme for IoT applications. IEEE Internet of Things Journal. 2020. vol. 7(4). pp. 3474–3484.
14. Chaabouni N., Mosbah M., Zemmari A., Sauvignac C., Faruki P. Network intrusion detection for IoT security based on learning techniques. IEEE Communications Surveys and Tutorials. 2019. vol. 21(3). pp. 2671–2701.
15. Aldweesh A., Derhab A., Emam A.Z. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems. 2020. vol. 189(5). DOI: 10.1016/j.knosys.2019.105124.
16. Albulayhi K., Sheldon F.T. An adaptive deep-ensemble anomaly-based intrusion detection system for the internet of things. In 2021 IEEE World AI IoT Congress (AIIoT). IEEE, 2021. pp. 0187–0196.
17. Alrubayyi H., Goteng G., Jaber M., Kelly J. Challenges of malware detection in the IoT and a review of artificial immune system approaches. Journal of Sensor and Actuator Networks. 2021. vol. 10(4). DOI: 10.3390/jsan10040061.
18. Al-Turaiki I., Altwaijry N. A convolutional neural network for improved anomaly-based network intrusion detection. Big Data. 2021. vol. 9(3). pp. 233–252.
19. Lam N.T. Detecting unauthorized network intrusion based on network traffic using behavior analysis techniques. International Journal of Advanced Computer Science and Applications. 2021. vol. 12(4). DOI: 10.14569/IJACSA.2021.0120407.
20. Aljumah A. IoT-based intrusion detection system using convolution neural networks. PeerJ Computer Science. 2021. vol. 7. DOI: 10.7717/peerj-cs.721.
21. Akhtar M.S., Feng T. Deep learning-based framework for the detection of cyberattack using feature engineering. Security and Communication Networks, 2021. no. 1. DOI: 10.1155/2021/6129210.
22. Liu C., Gu Z., Wang J. A hybrid intrusion detection system based on scalable K-means+ random forest and deep learning. IEEE Access. 2021. vol. 9. pp. 75729–75740.
23. Thilagam T., Aruna R. Intrusion detection for network based cloud computing by custom RC-NN and optimization. ICT Express. 2021. vol. 7(4). pp. 512–520.
24. Kanna P.R., Santhi P. Unified deep learning approach for efficient intrusion detection system using integrated spatial-temporal features. Knowledge-Based Systems. 2021. vol. 226. DOI: 10.1016/j.knosys.2021.107132.

25. Yin S.L., Zhang X.L., Liu S. Intrusion detection for capsule networks based on dual routing mechanism. Computer Networks. 2021. vol. 197. DOI: 10.1016/j.knosys.2021.107132.

26. Khan A.S., Ahmad Z., Abdullah J., Ahmad F. A spectrogram image-based network anomaly detection system using deep convolutional neural network. IEEE access. 2021. vol. 9. pp. 87079–87093.

27. Chen Y., Lin Q., Wei W., Ji J., Wong K.C., Coello C.A.C. Intrusion detection using multi-objective evolutionary convolutional neural network for Internet of Things in Fog computing. Knowledge-Based Systems. 2022. vol. 244. DOI: 10.1016/j.knosys.2022.108505.

28. Dahou A., Abd Elaziz M., Chelloug S.A., Awadallah M.A., Al-Betar M.A., Al-Qaness M.A., Forestiero A. 2022. Intrusion detection system for IoT based on deep learning and modified reptile search algorithm. Computational Intelligence and Neuroscience. 2022. no. 1. DOI: 10.1155/2022/6473507.

29. Haq M.A., Rahim Khan M.A., AL-Harbi T. Development of PCCNN-based network intrusion detection system for EDGE computing. Computers, Materials and Continua. 2022. vol. 71(1). DOI: 10.32604/cmc.2022.018708.

30. Albulayhi K., Abu Al-Haija Q., Alsuhibany S.A., Jillepalli A.A., Ashrafuzzaman M., Sheldon F.T. IoT intrusion detection using machine learning with a novel high performing feature selection method. Applied Sciences. 2022. vol. 12(10). DOI: 10.3390/app12105015.

31. Stoyanova M., Nikoloudakis Y., Panagiotakis S., Pallis E., Markakis E.K. A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Communications Surveys and Tutorials. 2020. vol. 22(2). pp. 1191–1221.

32. Henderi H., Wahyuningsih T., Rahwanto E. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. International Journal of Informatics and Information Systems. 2021. vol. 4(1). pp. 13–20.

33. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002. vol. 16. pp. 321–357.

34. Allam M., Nandhini M. Optimal feature selection using binary teaching learning based optimization algorithm. Journal of King Saud University-Computer and Information Sciences. 2022. vol. 34(2). pp. 329–341.

35. Smys S., Basar A., Wang H. Hybrid intrusion detection system for internet of things (IoT). Journal of ISMAC 2020. vol. 2(04). pp. 190–199.

36. Raichura M., Chothani N., Patel D. Efficient CNN-XGBoost technique for classification of power transformer internal faults against various abnormal conditions. IET Generation, Transmission and Distribution. 2021. vol. 15(5). pp. 972–985.

37. Ullah I., Mahmoud Q.H. Design and development of RNN anomaly detection model for IoT networks. IEEE Access. 2022. vol. 10. pp. 62722–62750.

38. Susilo B., Sari R.F. Intrusion detection in IoT networks using deep learning algorithm. Information. 2020. vol. 11(5). DOI: 10.3390/info11050279.

39. Soliman S., Oudah W., Aljuhani A. Deep learning-based intrusion detection approach for securing industrial Internet of Things. Alexandria Engineering Journal. 2023. vol. 81. pp. 371–383.

40. Khelil H., Brahimi M. Toward an efficient web service composition based on an improved BTLBO algorithm. The Journal of Supercomputing. 2024. vol. 80(7). pp. 8592–8613.

41. Ullah I., Mahmoud Q.H. A framework for anomaly detection in IoT networks using conditional generative adversarial networks. IEEE Access. 2021. vol. 9. pp. 165907–165931.
42. Khuat T.T., Le M.H. Binary teaching–learning-based optimization algorithm with a new update mechanism for sample subset optimization in software defect prediction. Soft Computing. 2019. vol. 23(20). pp. 9919–9935.
43. Nazir A., He J., Zhu N., Qureshi S.S., Qureshi S.U., Ullah F., Wajahat A., Pathan M.S. (2024). A deep learning-based novel hybrid CNN-LSTM architecture for efficient detection of threats in the IoT ecosystem. Ain Shams Engineering Journal. 2024. vol. 15. no. 7. DOI: 10.1016/j.asej.2024.102777.
44. Gao X., Jamil N., Ramli M.I., Ariffin S.M.Z.S.Z. A Comparative Analysis of Combination of CNN-Based Models with Ensemble Learning on Imbalanced Data. JOIV: International Journal on Informatics Visualization. 2024. vol. 8. no. 1. pp. 456–464.
45. Zawaideh F.H., Al-Asad G., Swaneh G., Batainah S., Bakkar H. Intrusion Detection System for (IoI) Networks Using Convolutional Neural Network (CNN) and Xgboost Algorithm. Journal of Theoretical and Applied Information Technology. 2024. vol. 102(4). pp. 1750–1759.
46. Swana E.F., Doorsamy W., Bokoro P. Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset. Sensors. 2022. vol. 22(9). DOI: 10.3390/s22093246.

Нараянарао Чоккапу — научный сотрудник, кафедра компьютерных наук и инженерии, Технологическая школа GITAM. Область научных интересов: безопасность интернета вещей, глубокое обучение. Число научных публикаций — 3. nchokkap@gitam.in; кампус GITAM Вишакхапатнам, Ганди Нагар, Рушиконда, 530045, Вишакхапатнам, Андхра-Прадеш, Индия; р.т.: +91(08912)790-501.

Мандапати Венкатесвара Рао — профессор, кафедра компьютерных наук и инженерии, Технологическая школа GITAM. Область научных интересов: робототехника, облачные вычисления. Число научных публикаций — 10. vmandapa@gitam.edu; кампус GITAM Вишакхапатнам, Ганди Нагар, Рушиконда, 530045, Вишакхапатнам, Андхра-Прадеш, Индия; р.т.: +91(08912)790-501.

Бодду Бхаскара Рао — доцент, кафедра компьютерных наук и инженерии, Технологическая школа GITAM. Область научных интересов: машинное обучение, наука о данных, семантическая сеть. Число научных публикаций — 8. bboddu@gitam.edu; кампус GITAM Вишакхапатнам, Ганди Нагар, Рушиконда, 530045, Вишакхапатнам, Андхра-Прадеш, Индия; р.т.: +91(08912)790-501.