Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

An Improved Hybrid Intrusion Detection Approach

Ankesh Gupta[1], Baldev Singh[2], Nilam Chaudhary[3]
[1,2]Dept. of CSE, VGU, Jaipur Rajasthan India
[3]Dept. of CSE, SKIT, Jaipur Rajasthan India

Abstract: **-** In today's modern era, our always-connected world is lined with smart devices which is a double-edged sword. Although it is tremendously convenient, on the other hand, our data and systems are progressively exposed to a growing army of hackers. In this regard, the need for a reliable IDS continues to grow. The reason is that intrusion detection systems is vital not only to protect data but to prevent computer systems from unauthorized access and cyber-threats. Conventional IDSs limitations lie in the dependence on the signature base detection, which renders them unable to identify unknown and harmful threats. In many ways, machine learning is a promising approach to the identification of such malicious activity through the development of efficient, high-performing solution. In this study, we propose a different hybrid approach that combines two of the most well-known methods of ML; the J48 and Random Forest. The proposed approach is also made more effective through the use of a recursive feature elimination process. This process selects the ideal subset of relevant features and increases the performance of the model. The proposed approach has been examined and tested using the NSL-KDD intrusion detection benchmark datasets, which covers all kinds of intrusions. The results demonstrate the ability of proposed method to effectively detect different types of intrusions and well compete with other state-of-the-art intrusion detection methodologies.

Keywords*: Cyber-attacks, Denial of Service (DoS), Probe, Remote to Local (R2L), User to Root (U2R), Decision Tree (DT), Random Forest (RF), Layered Framework*

1. Introduction

The explosion of internet connections and interconnected digital gadgets in modern society dramatically increase the threats of cyberattacks over the past decade and has made data and connected system security at a top priority [1]. As information processing and internet access become cheaper, more organizations are at risk from a wider range of cyber threats. This is why Intrusion Detection Systems (IDS) are becoming increasingly important for data and device security. However, various methods and strategies have been utilized in the past and will continue to be employed in the foreseeable future in the domain of IDS. The overarching desire for a network security solution that offers absolute safety and reliability is universal. However, despite extensive research and the availability of numerous options, attaining a loyal solution remains a challenge until all the prerequisites of the intrusion detection system are fulfilled. These prerequisites are not only time-consuming and complex but also require frequent updates to ensure effectiveness [2]. There are two main types of IDS [3]: misuse detection, which looks for known attack patterns, and anomaly detection, which identifies unusual activity. Although signature-based misuse detection is the most common method currently used, researchers are actively exploring how to apply advanced machine learning for intrusion detection. While signature-based IDS are effective but they face huge hitches to detect new types of attacks or variations of existing ones. This is also a problem for IDS that rely on supervised learning methods. In a study done by P. Laskov et al. published in 2006 [4], different methods of intrusion detection had been compared. They conclude that the one perfect method is unobtainable when dealing with the data that comes from a different context comparing to training data. The grouped method consisting of two approaches in combination might be a a more talented method for intrusion detection. Oumaima Chakir and her team [5] conduct comprehensive research on the benefits and obstacles of using a wide range of combined techniques. In their work, they deeply explore the effectiveness and possibility of these merged approaches in implementing measures on multiple issues in IDS. Similarly, Md.

Alamgir Hossain and his group [6] illustrate the indication of one of the ensemble methods' usefulness in intrusion detection. They, however, observe that a lot of scholars indicate in their paper that combining numerous advanced techniques can improve the performance of intrusion detection. This research encourages our study in order to implement improvements where necessary.

In this paper, a novel hybrid approach has been introduced which integrates two well-known machine learning techniques the decision tree J48 algorithm and Random Forest. It is used to demonstrate effective intrusion detection system. This fusion enables accurate and efficient detection possibility of various classes' attacks. A recursive feature elimination process is used to further reduce the features and jointly improve accuracy of intrusion detection by eliminating unimportant features and preserving important features through ranking. The experimental validation process uses a 10-fold cross-validation technique. The test results on NSLKDD dataset show that our approach can get good detection accuracy and it can be a real practical network anomaly detection system.

In summary, we have made the following main contributions in our work:

1. Take a look at the latest methods for detecting intrusions and analyse both their advantages and disadvantages.
2. Compare the performance of various machine learning algorithms in detecting new intrusions or attacks.
3. Propose an intrusion detection hybrid mechanism using both the J48 decision tree and the RF technique for improving the category-wise detection of intrusions.
4. A recursive feature removal method is employed to tackle feature redundancy. This approach efficiently screens the dataset to identify the most valuable features that positively impact the model's accuracy.
5. To showcase the efficacy of the suggested techniques, we carry out experiments on NSLKDD intrusion datasets. The outcomes of these experiments distinctly indicate that our proposed approach yields a higher detection rate and lower false positive rate compared to alternative methods.

The subsequent sections of this paper are meticulously organized to provide a comprehensive examination of the proposed approach. In Section 2, an in-depth review of the relevant literature surrounding Intrusion Detection Systems (IDS) is presented. Section 3 intricately details the complete detection framework and its associated methodology, providing readers with a clear understanding of the proposed approach. The empirical validation of the suggested methodology is rigorously demonstrated through comprehensive tests in Section 4. Finally, Section 5 encapsulates key findings, draws pertinent conclusions, and identifies avenues for further research.

2. Related Work

This section briefly sums up past research on using computers to catch intruders, like hackers, and talks about what works well and what doesn't. It leads into the main idea of our paper, which is a new approach using a mix of machine learning methods. Scientists have tried many ways to use machine learning techniques to spot intruders over time. We'll compare recent methods to understand why our new approach matters for finding intruders and keeping data and systems secure.

Al-Yaseen et al. [7] introduced a hybrid intrusion detection system that integrates five tiers of Extreme Learning Machine (ELM) and Support Vector Machine (SVM). Initially, traffic is categorized into four types: DoS, Probe, User to Root (U2R), and Remote to Local (R2L) attacks, with R2L and U2R considered least likely due to their similarity to regular connections. Each level uses a different classifier; levels 1, 3, 4, and 5 adapt four SVM classifiers, then Level 2 adopt ELM classifier for higher detection of the probe. After preprocessing the KDD dataset, they experimented modified K-means for feature extraction. Their model gave the result of 95.75% accurate performance, which is slightly better than multi-level SVM (95.57%), also the false alarm rate was 1.87% which is lesser than 2.17%. Papamartzivanos et al. [8] introduced an original intrusion detection approach, named Dendron, with the Genetic Algorithm in constructing the Decision Tree (DT) classifiers to detect abusive systems. State-of-the-art identification accuracies were reached by Dendron on 3 datasets: 99% on KDDCup'99 (average 89%), 97% on NSL-KDD (average 90%), and 52% on UNSWNB15 (average 84%). This version displays the heterogeneous performance of Dendron across datasets under specific hacking

conditions. Of course, a major drawback is that it can not be quickly used to pinpoint new attacks, severely limiting its application to data sets or types of attacks. Narayana Rao et al. [9] had used the methods of Autoencoder (AE) and Deep Neural Network (DNN) for attack classification, Chang et al. [10] had used the combination of Random Forest (RF) with Support Vector Machine (SVM) method intrusion detection. Though all the methods suffer from various problems, applying the techniques produces better output; one reason being that they are inefficient against unknown attacks that do not have a signature in the training data set. To cope with the emerging issues related to high-dimensional information and data imbalances, recent researchers have tried devising better classifier methodologies; notable among them are Di Mauro et al. [11], Abdulhammed R. et al. [12], and Seo, J.H. et al. [13]. In the quest to reduce the irrelevant features and enhance the detection accuracy, Zhou Y. et al. [14] had introduced an initial step of relevance testing for features, associating correlational analysis with the set of classified samples. They are followed by the adoption of the composite classifier method with a variety of methods to partition the dataset in an efficient manner. The methodology adopted, therefore, must involve a proactive commitment to the enhancement of quality and accuracy of classification systems in today's complex surrounding within the context of modern data landscapes. A number of studies carried out on this have observed how the capability of intrusion detection systems is augmented with machine learning [15-31]. In most cases, they found that a linkage with machine learning added to the strength and reliability of the currently established system. This means that the use of machine learning in detecting intrusions greatly strengthens data and system security.

In line with relevant field studies, the major goals of research on intrusion detection systems are to reduce false alarm and improve the accuracy of models and detection rate. In addition, the different ways the data should be processed, be it selection of the features or reduction in the dimensionality of the data, have been illustrated for probability to minimize the use of the computer resources and improve model resourcing. This work is focused on the design of a very important intrusion detection method, relevant for cybersecurity, with high detection rates and low false alarms. Unknown and minor class attack classification, like R2L or U2R, on the other hand, are part of the very important but very challenging task of the real deployment of a system for the detection of intrusion in effective and efficient measures. In this overview, we shall present a hybrid method of recursive feature reduction and Decision Trees (DT) and Random Forest (RF) for that purpose. The major goal is, hence, accuracy enhancement in IDS and resolution of the IDS accuracy dilemma.

3. Architecture of Proposed Practice

The figure 1 well defines the architecture of the proposed hybrid methodology that contains five integrated stages: Input data acquisition, preprocessing procedures, feature extraction methods, classification algorithms, and finally, the generation of final output. This structure would help to have a clear notion about the operation mechanism of the proposed hybrid technique. The proposed method bears high focus on the complete detection of all these four different categories of cyber attacks stated as: Denial of Service (DoS), Probe (Prob), Remote-to-Local (R2L), and User-to-Root (U2R). This is achieved by the use of Decision Trees (DT) and Random Forest (RF) techniques that are part of the layered framework with a constrained feature set; this primarily aims to minimize false alarms and ensure the system has high accuracy in anomaly detection and other malicious attacks.

The proposed attack detection system, depicted in Figure 1, is based on a multi-level approach, where the incoming packets are analyzed at four hierarchical levels to identify potential events of attacks. At each level, packets flagged as potential attacks are promptly blocked, while those not identified as such proceed to the subsequent level for evaluation against different attack types. Each stage, except the final level, follows a uniform procedure for event validation and forwards the packet to the next level for further analysis. In cases where an incoming packet is not recognized as an attack event at any level, it is considered normal data and permitted to undergo standard processing within the system. This hierarchical strategy secures the model's overall accuracy while notably enhancing its precision, especially for minor class attacks like U2R and R2L.

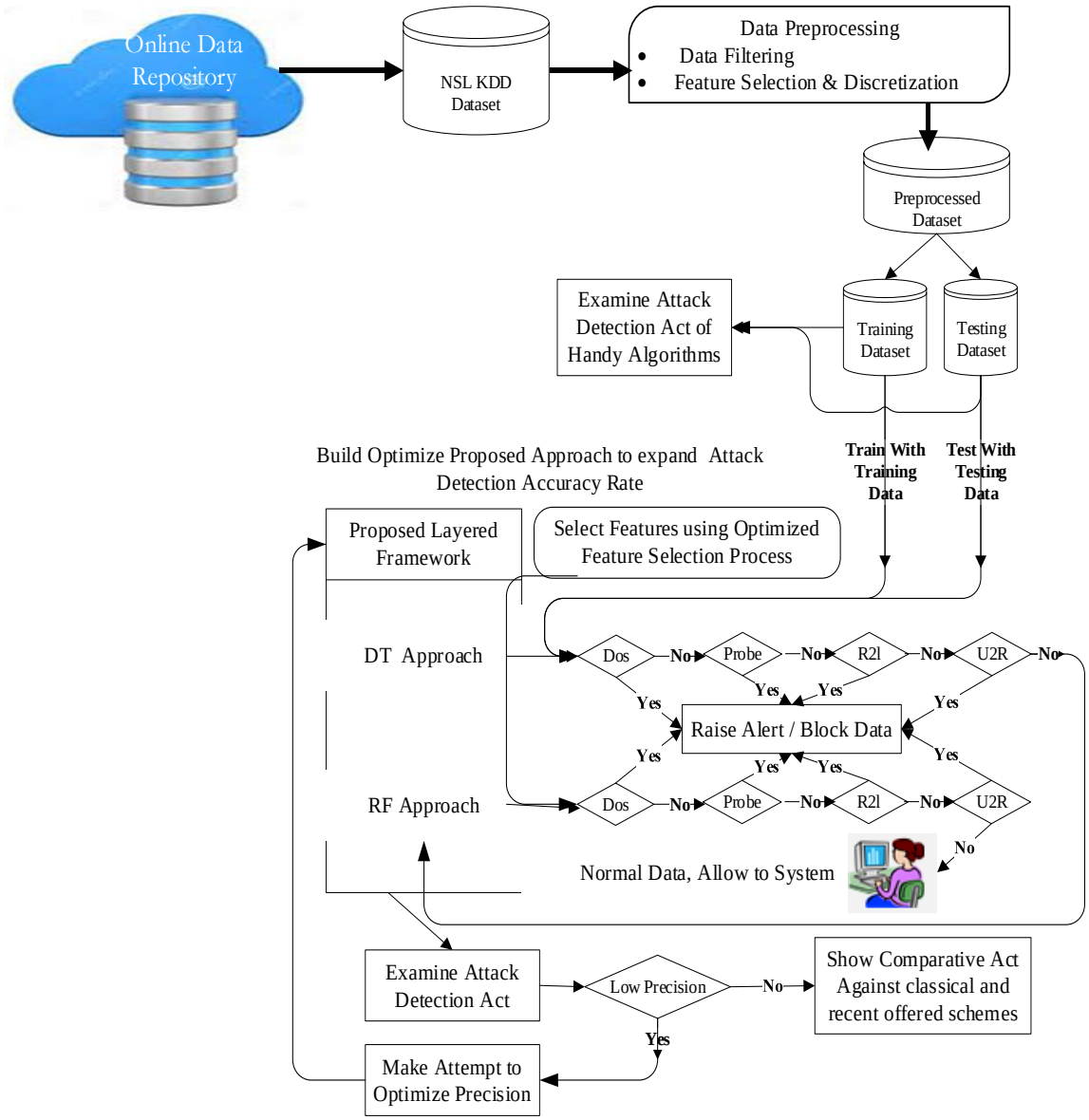

Fig. 1: Proposed Approach for Enhance Detection Accuracy of Intrusion

3.1 Dataset Description

As integral components within machine learning frameworks, datasets furnish vital information for learning processes and facilitate precise predictions or assessments. Moreover, a superior dataset containing precise and relevant instances catalyses the refinement of a more dependable and precise model. In this investigation, we utilize the NSLKDD dataset [34], an optimized version derived from the well-known KDD-CUP99 [35] dataset.

The effective rectification of redundant attributes and replicated entries within the conventional KDD-CUP99 dataset has propelled its prominence in the realm of network security intrusion detection. Furthermore, the NSL-KDD dataset's classification of network connections into normal and problematic categories significantly augments its efficacy within this field. Following figure denoting the attack signatures of NSLKDD dataset.

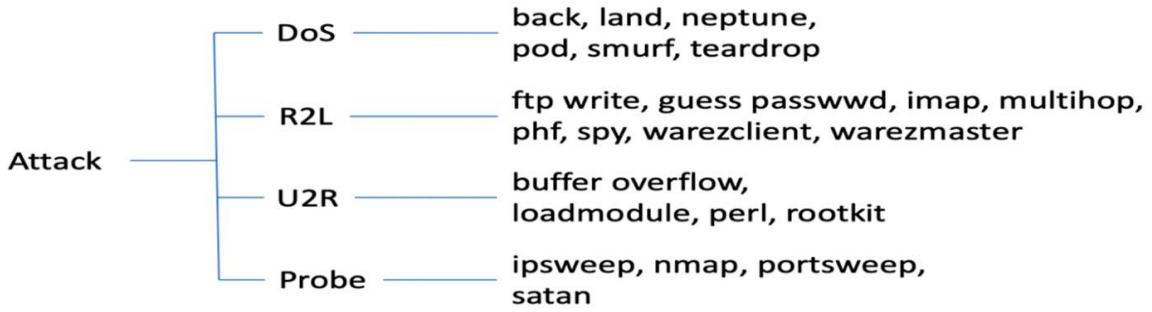

Fig. 2: Attack Category Associated with NSLKDD Dataset

3.2 Data preprocessing

The data preprocessing stage plays a pivotal role in refining, converting, and preparing raw data for analysis and modelling. It ensures that the data is in a suitable format and quality for processing by machine learning algorithms. Common tasks during data preprocessing involve handling missing values, eliminating outliers, standardizing or normalizing data, and encoding categorical variables. By implementing these procedures, data preprocessing significantly enhances the accuracy, effectiveness, and efficiency of subsequent machine learning models, contributing to improved outcomes [36]. We do strongly affirm that within the NSLKDD dataset, there exists no duplication of data. But the data input in most machine learning methods is supposed to be of numeric format. The said classification method is actually designed for work using numeric data. We go on to convert categorical data into numeric form through a defined method regarding data conversion, known as one-hot encoding. This ensures that machine learning techniques work efficiently within the datase, facilitating continuous and precise analysis of the data.

3.3 Feature selection

Feature selection plays a key process in developing the machine-learning model for identifying network intrusion. Since there is a probability that not all features input into the dataset may play an important and equal role in the prediction for network intrusion, it probably overleads to overfitting the model with data when complicated with unnecessary or duplicate characteristics [37]. In order to cope with this, our method employs recursive feature elimination. This will be done through the cyclic evaluation and pruning of systematically less informative features in order to afford better effectiveness. Hence, at the same time, the model finds network intrusion but not to the point of overfitting the data.

Pseudo code of optimized feature selection procedure:

1. Initialize an empty list X to store features of NSL KDD dataset.
2. Populate list X with associated features of NSL KDD dataset.
3. Analyse the performance of the chosen scheme using the entire feature set of X:
   - Set PR as the precision value of the attack detection.
4. Initialize a variable J to represent the highest-ranked feature index.
5. Iterate through each feature in X:
   a. Remove feature Xj from list X.
   b. Evaluate the performance of the chosen scheme without feature Xj:
      - Calculate the precision value of the attack detection, denoted as NPR.
   c. Compare NPR with PR:
      - If NPR > PR:
         - Update PR to NPR.

- Continue to the next iteration.

- If NPR <= PR:

- Re-insert feature Xj into list X.

- Increment J by 1.

6. Repeat steps 5 until J is less than the length of X.

7. At the end of the iterative process, X will contain only the set of optimal features recommended for enhancing the attack detection performance of the chosen scheme.

3.4 Hybrid Process for IDS

By merging the strategies of intrusion detection with additional functionalities, the hybrid intrusion detection method enhances both accuracy and durability in identifying potentially harmful activities within computer networks. This amalgamation of multiple detection techniques ensures comprehensive network security while concurrently boosting detection rates and fortifying resilience against a diverse array of threats. Through the integration of various detection methods, this hybrid approach optimizes the network defines mechanism, enhancing its ability to discern and mitigate potential security breaches effectively. In the proposed model DT and RF techniques deployed in a layered form to fulfil the intended objectives within the domain of intrusion detection. A very short description of those is given below:

3.4.1 Decision Tree (J48)

Decision Trees are one of the best tools to work with large datasets. This method comes as one with a model based on decisions. It uses the tree structure in its design. In this model, we encode the leaves for a particular class, while the inner nodes are all decision nodes. This is a process whereby a complex problem will be broken down into smaller subproblems until it eventually gets solved. One of the nodes in the tree encodes a feature of the problem in order for it to drive the traversal through its importance. As a matter of fact, according to N. Bhargava et al. [33], Decision Trees represent structured paths of decision-making techniques and, therefore, offer a systematic way for analysis and solution to complex decision situations, which are very appropriate for application in many academic disciplines.

This strategy has several advantages that come with it and can be realized, including

- A proficient approach for handling both numerical and categorical data.
- Easily understandable.
- Capable of identifying the best, average, and worst values for different scenarios.
- Achieves precise and rapid classification even in cases of redundant attributes or unidentified records, effectively addressing issues arising from multiple outputs.
- Unlike current practices, this strategy eliminates the requirement for preprocessing steps such as normalization and removal of blank values.
- Can employ standard statistical tests to assess the reliability of the model.

Some of the key Limitation of Decision Tree are

- Significantly diverse decision trees may be generated from slight differences in the input data. The consistency and dependability of the intrusion detection system may be impacted by this instability.
- Decision trees may generate biased trees that are more suited to the dominant class when specific attack types predominate in the dataset. For intrusion types that happen less frequently, this may result in less-than-ideal detection performance.
- Since many conventional decision tree techniques are made to handle discrete values, managing continuous

variables may call for additional steps in the preprocessing stage or the application of particular algorithms.

- When decision trees encounter irrelevant features in the dataset, they may have trouble formulating decisions based on them, which could produce less-than-ideal outcomes.
- Detecting specific sorts of intrusions may be hindered by decision trees', poor recital in the presence of complicated and sophisticated structures in the dataset.

3.4.2 Random Forest (RF)

Bagging and feature randomness are combined in the Random Forest (RF) classifier, which is described in depth by Breiman, L. [32]. In order to aid in classification, it trains many Decision Trees (DTs). If-then rules are created using Decision Trees, which divide the feature space recursively. In order to improve prediction performance and reduce overfitting, RF excels in a variety of DTs. Whereas feature randomness chooses random feature subsets for each node, bagging trains DTs on random subsets of input. Application-wide, RF strikes a balance between precision and complexity.

Some of key advantages of Random Forest for Intrusion Detection System are:

- By combining several decision trees, the ensemble approach Random Forest improves intrusion detection accuracy and generalization.
- Compared to individual decision trees, Random Forest is less prone to overfitting, which improves its capacity to adapt well to new data.
- It helps in identifying important characteristics that contribute to intrusion detection by providing a measure of feature relevance.
- Suitable for various types of intrusion detection tasks, including binary and multiclass classification.

Some of Limitations of Random Forest Approach for Intrusion Detection System are:

- Because of its intricacy, the Random Forest model can be difficult to analyse and comprehend, which results in a "black-box" style of decision-making.
- It can take a lot of processing power to train a random forest, especially when there are many trees or deep trees. This could be a drawback in settings with limited resources.
- Random Forest is known to use a lot of memory, especially when processing a lot of decision trees.
- While Random Forest is generally robust to overfitting, it can still be susceptible to overfitting in the presence of noisy or irrelevant features in the dataset.

4. Evaluation Setup and Result Discussion

In this segment, we present the effectiveness of our proposed intrusion detection model, assessed on a computing platform comprising a Core-i7 13700K CPU@ 2.50 GHz and 96 GB of DDR4 memory, operating on the Windows 10 Professional 64-bit platform. Feature selection and model training were conducted using Eclipse Juno version 2023-09. To validate our model's performance, we adopted the widely recognized 10-fold cross-validation method, partitioning the dataset into ten subsets for training and testing purposes. Through ten iterations of this process, we rigorously evaluated the model's capabilities, particularly its predictive accuracy in detecting intrusions, using the confusion matrix (Table 1). By leveraging an array of evaluation metrics, we have substantiated the effectiveness of our proposed approach in identifying and mitigating intrusion attempts, thereby empowering us to refine and optimize its precision and efficacy further.

Table 1: Confusion Matrix

	Predicted class	
	Attack	Normal

Actual class	Attack	TP	FN
	Normal	FP	TN

True positive (TP): The traffic is an attack and is correctly classified as attack traffic by the model.

True Negative (TN): The traffic is normal data and is correctly classified as normal by the model.

False positive (FP): The traffic is normal data and is classified as attack data by the model.

False Negative (FN): The traffic is an attack data and is classified as normal data by the model.

The approach is evaluated using the performance measures listed below.

$$Detection\ Accuracy = (TP)/(TP + TN) \quad (1)$$

$$FalsePossitiveRate = FP/(FP + TN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

To showcase the efficiency of the suggested method, it has been compared against commonly used algorithms including Naïve Bayes (NB), Hoeffding Tree (HT), J48, and a standalone version of the Random Forest algorithm. Following table and figures presents the outcomes of evaluation process for reference.

Table 2: Attack Detection Performance of Proposed and Classical IDS Technique

Techniques	Detection Accuracy (%) of Each Attack Type			
	DoS	Probe	R2L	U2R
NB	96.08	98.24	94.27	40.38
J48	99.94	99.38	94.17	51.92
RF	99.98	99.75	96.08	53.84
HT	99.07	93.29	91.05	25
Proposed	99.97	99.47	97.19	80.77

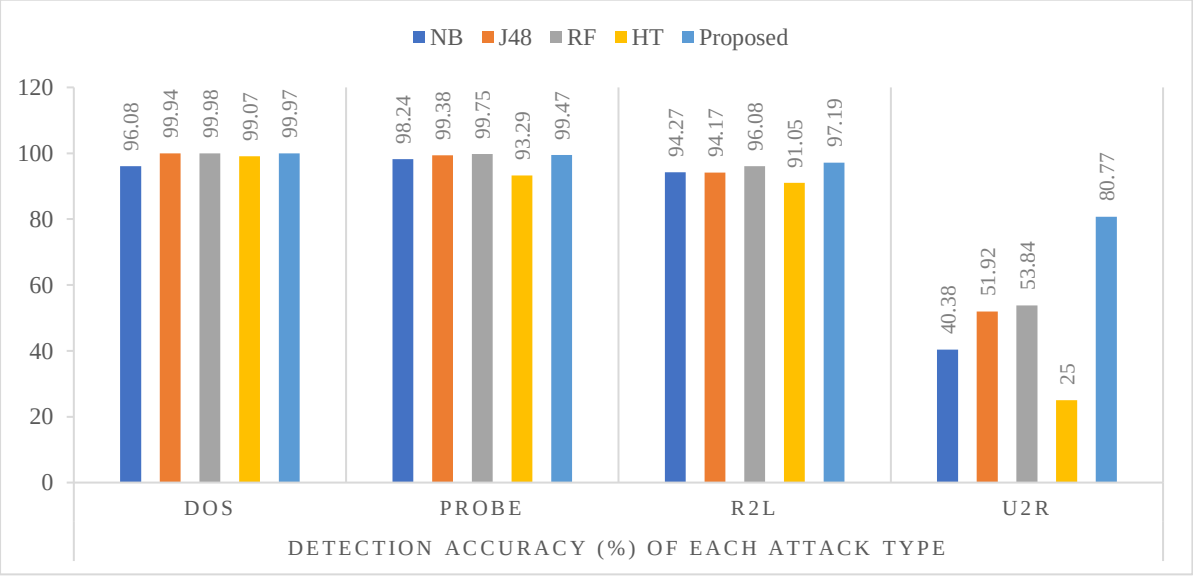

Fig. 3: Intrusion Detection Accuracy on NSLKDD Dataset

Table 3: FPR of Proposed and Classical IDS Technique

Techniques	False Positive Rate of Each Attack Type			
	DoS	Probe	R2L	U2R
NB	0.021	0.009	0.046	0.56
J48	0.048	0.006	0.055	0.42
RF	0.015	0.002	0.035	0.54
HT	0.006	0.053	0.069	0.75
Proposed	0.004	0.002	0.028	0.19

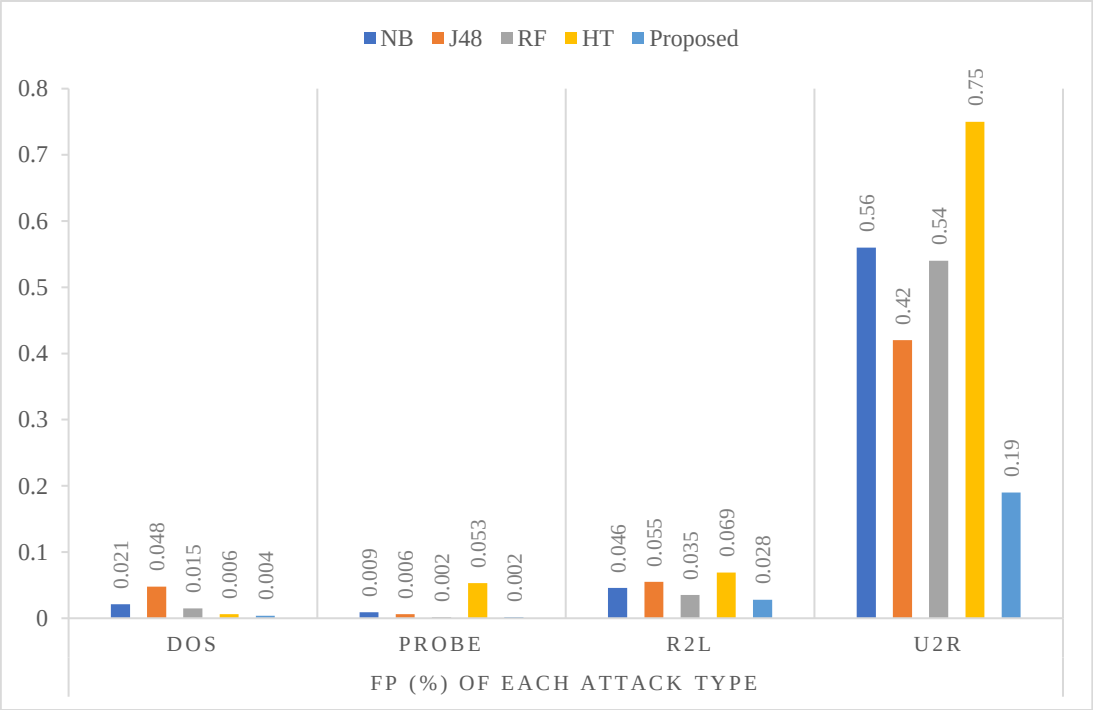

Figure 4: Comparative False Positive Rate

Table 4: Precision of Proposed and Classical IDS Technique

Techniques	Precision Rate			
	DoS	Probe	R2L	U2R
NB	99.5	89.0	89.6	8.8
J48	99.9	99.4	97.3	90
RF	1.00	99.8	96.4	53.8
HT	99.4	97.6	53.0	20.3
Proposed	99.93	99.92	98.77	90.38

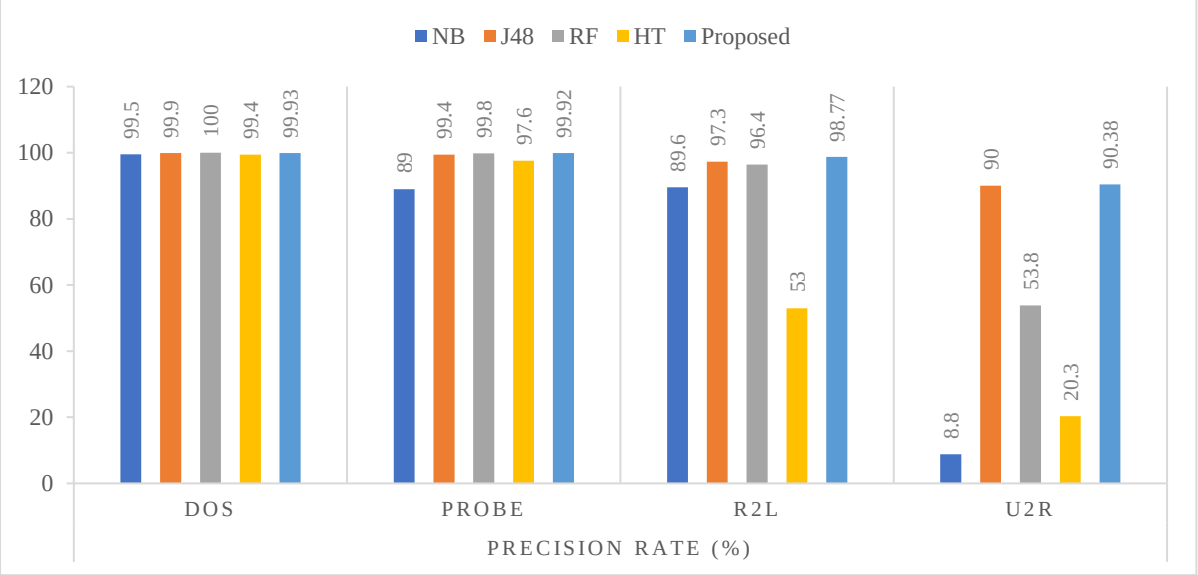

Fig. 5: Comparative Precision of Proposed and Classical IDS Technique

The comprehensive evaluation of conventional methods has revealed their effectiveness in detecting major class attacks such as DoS and Probe, yet they fall short in identifying minor class attacks within an acceptable threshold. The proposed approach overcome this inefficiency in an effective manner as shown in comparative fallouts in above tables and figures. The efficiency of proposed approach has also compared with the some of modern alternative offered techniques for detection of intrusion. Table 5 presents a visual depiction of the comparative analysis between the proposed technique and alternative methods.

Table 5 Comparative ID Accuracy Among Proposed and Alternative Methods.

Author(s)	Attained Accuracy (%)
Proposed Hybrid	99.82
Das T et. al., 2023 [38]	99.00
S. Mohamed and R. Ejbali, 2023, [39]	84.36
S. M. Kasongo, 2023 [40]	85.93
Reddy, G. et. al. , 2022 [41]	98.00
M. Rani, 2022 [42]	85.56
E. Mushtaq, 2022 [43]	79.00
W. L. Al-Yaseen et. al. , 2022 [44]	80.15
S. P. Thirimanne et. al. , 2022 [45]	81.87
Fu, Y et. al. , 2022 [46]	90.73
Nguyen Gia Bach et. al. , 2021 [47]	98.83
Al-Turaiki et. al., 2020 [48]	83.00
Elmasry W. et. al., 2020 [49]	96.91
Jiang K. et. al. 2020 [50]	83.58
Yang H. et. al., 2019, [51]	97.45

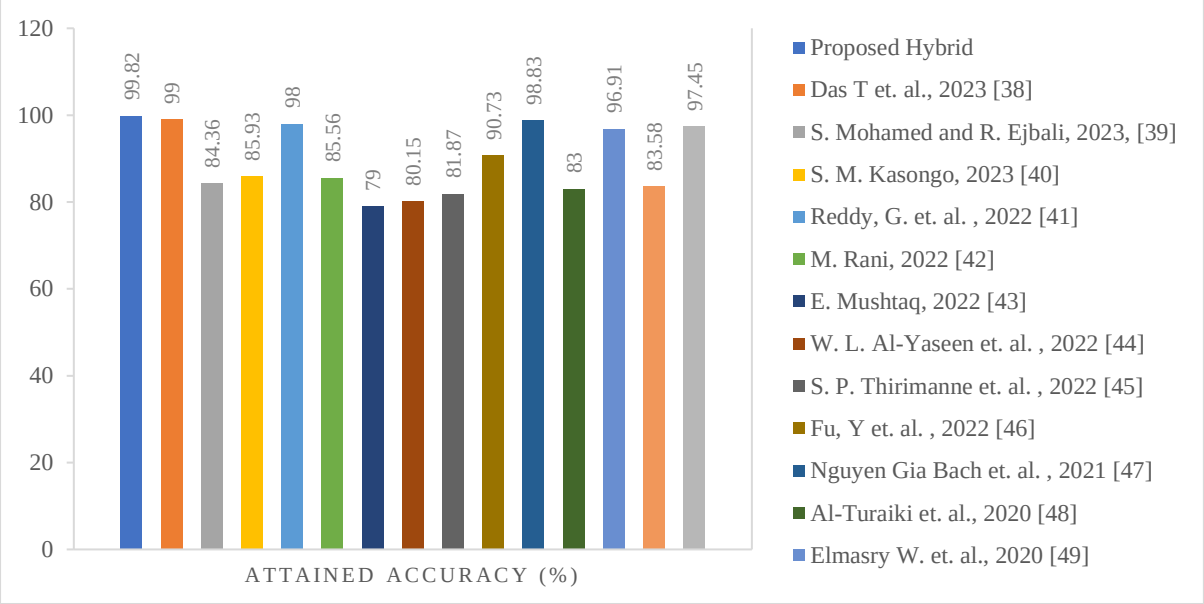

Fig. 6: Comparative Accuracy among Proposed and Modern Alternative IDS Method

The empirical evidence presented in the above tables and figures illustrates the enhanced accuracy of the proposed model over traditional as well as some modern offered techniques. Table 2 and Figure 3 elucidate that the Detection Rate (DR) of the proposed Intrusion Detection System (IDS) surpasses that of conventional methods when tested on the NSL-KDD dataset. Furthermore, the findings showcased in Tables 3 and Figure 4 indicate that our model exhibits significant improvements in reducing false alarm rates compared to established methods. For the NSL-KDD dataset, our proposed model achieves intrusion detection accuracies of 99.97% for DoS, 99.47% for Probe, 97.19% for R2L, and 80.77% for U2R attacks. While the Random Forest (RF) technique marginally enhances detection accuracy for DoS and Probe intrusions, our proposed approach significantly outperforms it in detecting more severe types of attacks, such as R2L and U2R. Table 5 and Figure 6 show the comparative attained accuracy among our proposed approach and some modern offered methods of intrusion detection. Comparative values significantly denote that the proposed approach is more suitable in comparison to other comparative solutions.

5. Conclusions

In the realm of network security, the significance of intrusion detection systems (IDSs) cannot be overstated, particularly in light of the escalating volume of network threats and technological advancements. IDSs have garnered substantial attention due to their pivotal role in bolstering network security. Our study focused on evaluating IDS efficacy using NSLKDD datasets, wherein a hybrid intrusion detection model integrating Decision Tree and Random Forest techniques exhibited superior performance compared to alternative strategies. The outcomes showcased remarkable accuracy and detection rates. The findings underscore the capability of the proposed approach to proficiently identify various attack types, thus serving as a valuable asset in fortifying the security of computer systems and networks against emerging cyber threats. In essence, our proposed methodology presents promising results, poised to contribute significantly to the development of more robust intrusion detection systems tailored for network security needs.

Refrences

[1] Successful Real-Time Security Monitoring, Riptech Inc. white paper, Sep. 2001.
[2] B. Yogesh, Dr. G. Suresh Reddy(2022), " Intrusion detection System using Random Forest Approach ", Turkish Journal of Computer and Mathematics Education, Vol.13, No.02, pp.730-731
[3] P. Animesh, J. Park, An overview of anomaly detection techniques: Existing solutions and latest technological trends, Comput. Networks, 51 (2007), 3448–3470.

[4] P. Laskov, P. Dussel, C. Schafer, et al., "Learning Intrusion Detection: Supervised or Unsupervised?" In Proc. of Image Analysis and Processing - ICIAP 2005, 13th International Conference, pp. 50-57, 2005.
[5] Oumaima Chakir, Abdeslam Rehaimi, Yassine Sadqi, El Arbi Abdellaoui Alaoui, Moez Krichen, Gurjot Singh Gaba, Andrei Gurtov, An empirical assessment of ensemble methods and traditional machine learning techniques for web-based attack detection in industry 5.0, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 3, 2023, Pages 103-119.
[6] Md. Alamgir Hossain, Md. Saiful Islam, Ensuring network security with a robust intrusion detection system using ensemble-based machine learning, Array, Volume 19, 2023, 100306.
[7] Al-Yaseen,W.L.; Othman, Z.A.; Nazri, M.Z.A. Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Syst. Appl. 2017, 67, 296–303.
[8] Papamartzivanos D, G´omez M´armol F, Kambourakis G. Dendron : Genetic trees driven rule induction for network intrusion detection systems. Future Generat Comput Syst Feb. 2018;79:558–74.
[9] Narayana Rao, K.; Venkata Rao, K.; P.V.G.D., P.R. A hybrid Intrusion Detection System based on Sparse autoencoder and Deep Neural Network. Comput. Commun. 2021, 180, 77–88.
[10] Chang, Y.; Li, W.; Yang, Z. Network intrusion detection based on random forest and support vector machine. In Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, China, 21–24 July 2017; Volume 1, pp. 635–638.
[11] Di Mauro, M.; Galatro, G.; Fortino, G.; Liotta, A. Supervised feature selection techniques in network intrusion detection: A critical review. Eng. Appl. Artif. Intell. 2021, 101, 104216.
[12] Abdulhammed, R.; Musafer, H.; Alessa, A.; Faezipour, M.; Abuzneid, A. Features dimensionality reduction approaches for machine learning based network intrusion detection. Electronics 2019, 8, 322.
[13] Seo, J.H.; Kim, Y.H. Machine-Learning Approach to Optimize SMOTE Ratio in Class Imbalance Dataset for Intrusion Detection. Comput. Intell. Neurosci. 2018, 2018, 9704672.
[14] Zhou, Y.; Cheng, G.; Jiang, S.; Dai, M. Building an efficient intrusion detection system based on feature selection and ensemble classifier. Comput. Netw. 2020, 174, 107247.
[15] E. Kabir, J. Hu, H. Wang and G. Zhuo, "A novel statistical technique for intrusion detection systems" Future Generation Computer Systems, vol. 79, pp. 303–318, 2018.
[16] H. Wang, G. Jie. and W. Shanshan, "An effective intrusion detection framework based on SVM with feature augmentation," Knowledge-Based Systems, vol. 136, pp. 130–139, 2017.
[17] N. Farnaaz and M. A. Jabbar, "Random forest modelling for network intrusion detection system" Computer Science, vol. 89, no. 1, pp. 213–217, 2016.
[18] M. Swarnkar and N. Hubballi, "OCPAD: One class naive bayes classifier for payload-based anomaly detection," Expert Systems with Applications, vol. 64, pp. 330–339, 2016.
[19] L. Yang, C. Ding, M. Wu and K. Wang, "Robust detection of false data injection attacks for data aggregation in an internet of things-based environmental surveillance," Computer Networks, vol. 129, pp. 410–428, 2017.
[20] C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," Computers & Security, vol. 70, pp. 255–277, 2017.
[21] G. Caminero, M. Lopez-Martin and B. Carro, "Adversarial environment reinforcement learning algorithm for intrusion detection," Computer Networks, vol. 159, pp. 96–109, 2019.
[22] J. Liu, J. He, W. Zhang, T. Ma, Z. Tang et al., "Anid-SEoKELM: Adaptive network intrusion detection based on selective ensemble of kernel elms with random features," Knowledge-Based Systems, vol. 177, pp. 104–116, 2019.
[23] F. Salo, A. B. Nassif and A. Essex, "Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection," Computer Networks, vol. 148, pp. 164–175, 2019.
[24] T. H. Divyasree and K. K. Sherly, "A network intrusion detection system based on ensemble CVM using ef□cient feature selection approach," Computer Science, vol. 143, pp. 442–449, 2018.

[25] O. Y. Al-Jarrah, Y. Al-Hammdi, P. D. Yoo, S. Muhaidat and M. Al-Qutayri, “Semi-supervised multilayered clustering model for intrusion detection,” Digital Communications and Networks, vol. 4, no. 4, pp. 277–286, 2018.
[26] A. A. Hady, A. Ghubaish, T. Salman, D. Unal and R. Jain, “Intrusion detection system for healthcare systems using medical and network data: A comparison study,” IEEE Access, vol. 8, pp. 106576–106584, 2020.
[27] S. Gao and G. Thamilarasu, “Machine-learning classi□ers for security in connected medical evices,” in Proc. ICCCN, Vancouver, Canada, pp. 1–5, 2017.
[28] I. Alrashdi, A. Alqazzaz, E. Alou□, R. Alharthi, M. Zohdy et al., “Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning,” in Proc. CCWC, University of Nevada, Las Vegas, NV, USA, pp. 305–310, 2019.
[29] A. F. Oliva, F. M. Perez, J. V. Berna-Martinez and M. A. Ortega, “Non-deterministic outlier detection method based on the variable precision rough set model,” Computer Systems Science and Engineering, vol. 34, no. 3, pp. 131–144, 2019.
[30] M. B. Nejad and M. E. Shiri, “A new enhanced learning approach to automatic image classification based on salp swarm algorithm,” Computer Systems Science and Engineering, vol. 34, no. 2, pp. 91–100, 2019.
[31] S. K. Singh, M. M. Salim, J. Cha, Y. Pan and J. H. Park, “Machine learning-based network sub-slicing framework in a sustainable 5G environment,” Sustainability, vol. 12, no. 15, pp. 1–22, 2020.
[32] Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32.
[33] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria, \Decision tree analysis on j48 algorithm for data mining," Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 6, 2013.
[34] Dua D, Graff C. KDD cup 1999 data data set. Univ. Calif. Irvine Sch. Inf. Comput. Sci.; 2019. Accessed: Jan. 25, 2023. [Online]. Available: http://archive.ics.uci.edu /ml.
[35] NSL-KDD dataset. Accessed: Feb. 05, 2023, https://www.unb.ca/cic/datasets/nsl. html.
[36] Subasi A. Data preprocessing. In: Practical machine learning for data analysis using Python. Elsevier; 2020. p. 27–89.
[37] Di Mauro M, Galatro G, Fortino G, Liotta A. Supervised feature selection techniques in network intrusion detection: a critical review. Eng Appl Artif Intell May 2021; 101:104216.
[38] Das T, Hamdan OA, Shukla RM, Sengupta S, Arslan E. UNR-IDD: intrusion detection dataset using network port statistics. In: 2023 IEEE 20th consumer Communications & networking conference (CCNC). Las Vegas, NV, USA: IEEE; Jan. 2023. p. 497–500.
[39] S. Mohamed and R. Ejbali, ‘‘Deep SARSA-based reinforcement learning approach for anomaly network intrusion detection system,’’ Int. J. Inf. Secur., vol. 22, no. 1, pp. 235–247, Feb. 2023.
[40] S. M. Kasongo, ‘‘A deep learning technique for intrusion detection system using a recurrent neural networks based framework,’’ Comput. Commun., vol. 199, pp. 113–125, Feb. 2023.
[41] Reddy, G.Vinoda & Kadiyala, Sreedevi & Potluri, Chandra & Saravanan, Shanthi & Athisha, G. & M a, Mukunthan & Sujaritha, M.. (2022). An Intrusion Detection Using Machine Learning Algorithm Multi-Layer Perceptron (MlP): A Classification Enhancement in Wireless Sensor Network (WSN). International Journal on Recent and Innovation Trends in Computing and Communication. 10. 139-145.
[42] M. Rani, ‘‘Effective network intrusion detection by addressing class imbalance with deep neural networks multimedia tools and applications,’’ Multimedia Tools Appl., vol. 81, no. 6, pp. 8499–8518, Mar. 2022.
[43] E. Mushtaq, A. Zameer, M. Umer, and A. A. Abbasi, ‘‘A two-stage intrusion detection system with auto-encoder and LSTMs,’’ Appl. Soft Comput., vol. 121, May 2022, Art. no. 108768.
[44] W. L. Al-Yaseen, A. K. Idrees, and F. H. Almasoudy, ‘‘Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system,’’ Pattern Recognit., vol. 132, Dec. 2022, Art. no. 108912.
[45] S. P. Thirimanne, L. Jayawardana, L. Yasakethu, P. Liyanaarachchi, and C. Hewage, ‘‘Deep neural network based real-time intrusion detection system,’’ Social Netw. Comput. Sci., vol. 3, no. 2, p. 145, Mar. 2022.

[46] Fu, Y.; Du, Y.; Cao, Z.; Li, Q.; Xiang,W. A Deep Learning Model for Network Intrusion Detection with Imbalanced Data. Electronics 2022, 11, 898. https://doi.org/10.3390/ electronics11060898

[47] Nguyen Gia Bach, Le Huy Hoang, and Tran Hoang Hai, "Improvement of K-nearest Neighbors (KNN) Algorithm for Network Intrusion Detection Using Shannon-Entropy," Journal of Communications vol. 16, no. 8, pp. 347-354, August 2021.

[48] Al-Turaiki I, Altwaijry N, Agil A, Aljodhi H, Alharbi S, Lina A. Anomaly-based network intrusion detection using bidirectional long short term memory and convolutional neural network. ISC Intl J. Inf. Secur. Nov. 2020;12(3):37–44.

[49] Elmasry W, Akbulut A, Zaim AH. Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. Comput Network Feb. 2020; 168:107042.

[50] Jiang K, Wang W, Wang A, Wu H. Network intrusion detection combined hybrid sampling with deep hierarchical network. IEEE Access 2020;8:32464–76.

[51] Yang H, Qin G, Ye L. Combined wireless network intrusion detection model based on deep learning. IEEE Access 2019;7:82624–32.