Neural Computing and Applications (2022) 34:15387–15395
https://doi.org/10.1007/s00521-020-04986-5

ORIGINAL ARTICLE

A stacked ensemble learning model for intrusion detection in wireless network

Hariharan Rajadurai[1] · Usha Devi Gandhi[1]

Received: 21 January 2020 / Accepted: 2 May 2020 / Published online: 21 May 2020

Abstract
Intrusion detection pretended to be a major technique for revealing the attacks and guarantee the security on the network. As the data increases tremendously every year on the Internet, a single algorithm is not sufficient for the network security. Because, deploying a single learning approach may suffer from statistical, computational and representational issues. To eliminate these issues, this paper combines multiple machine learning algorithms called stacked ensemble learning, to detect the attacks in a better manner than conventional learning, where a single algorithm is used to identify the attacks. The stacked ensemble system has been taken the benchmark data set, NSL-KDD, to compare its performance with other popular machine learning algorithms such as ANN, CART, random forest, SVM and other machine learning methods proposed by researchers. The experimental results show that stacked ensemble learning is a proper technique for classifying attacks than other existing methods. And also, the proposed system shows better accuracy compare to other intrusion detection models.

Keywords Network intrusion detection · Gradient boosting · Classification algorithms · Machine learning · Ensemble learning · Random forest tree

1 Introduction

At present, network security has become an important focus of computer security research. The attack on the network infrastructure is the threat against network and information security. An intrusion detection system (IDS) activities monitor and analysis user and system activity, system configuration auditing and ensuring critical system securities. The characteristics of effective intrusion detection systems are high detection rate, less false alarm, fewer CPU cycles and quick detection of intrusion. In [1], the IDS was developed by Denning in 1987. This paper was considered as a great statistics landmark in intrusion detection system (IDS) field. Every ID system must possess four characteristics. They are time, performance, dynamic reconfiguration and prediction performance. The IDS policy gets the requirements from the goal of IDS. The goals of the IDS involve enforcement of use policies, collection of evidence, detection and prevention of attacks. IDS classification techniques are shown in Fig. 1. IDS can also be classified into the following four categories on the basis of the detection approaches:

- Signature-based intrusion detection or misuse (knowledge-based)

In this approach, the comparison of the user's activity and known intrusion detection pattern is called signatures. The main advantage of signature-based intrusion detection is very easy to understand and high detection speed because of dealing with false positives.

- Anomaly-based intrusion detection system or behavior system

In this approach, normal user is compared with a new user to judge the new user's activity. The advantages of this method are less dependent on system software and very efficient to detect unknown and attacks.

- Host-based IDS (HIDS): HIDS is a method, which monitors the system activities are generally used for

✉ Usha Devi Gandhi
ushadevi.g@vit.ac.in

[1] School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, India

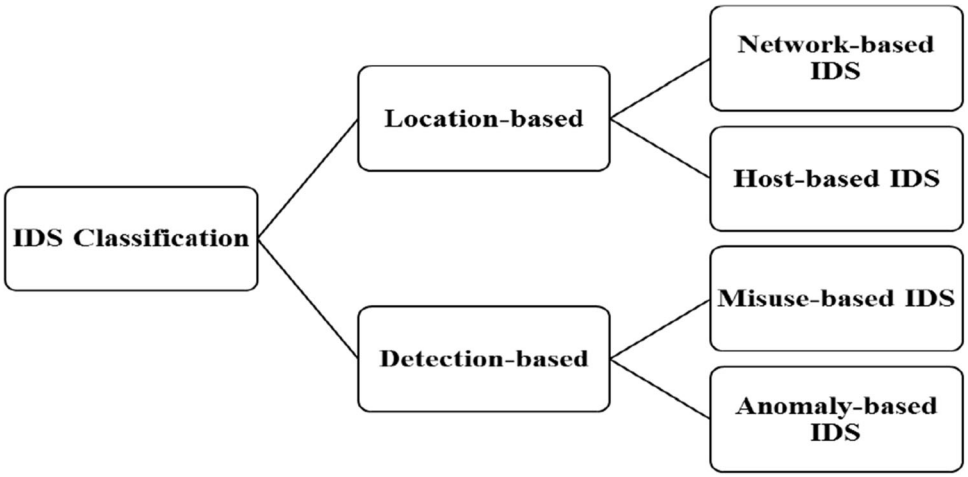


Fig. 1 Intrusion detection classification techniques

information collected by system logs and system audit trails.

- Network-based IDS (NIDS): NIDS is a system, it runs at strategic points in the network (e.g., server and switch).

In this paper, we propose stacked ensemble learning model for network intrusion detection by using gradient boosting machine (GBM) and random forest (RF) algorithms. The proposed method integrates the features of both GBM and RF classifiers. RF is the ensemble regression and classification approach. RF was developed by Breiman [2]. Random forest generates various classification trees. The primary features of the random forest algorithm are handling unbalanced data sets, running efficiently at the large data sets with various features and there is no nominal data problem. Gradient boosting stands for gradient descent and boosting is called by gradient boosting [3]. It is a powerful machine learning algorithm. General boosting algorithm works with a variety of loss functions. The models included in GBM are resistant regression, K-class classification, regression and risk modeling. Gradient boosting can do regression, classification and ranking.

2 Related work

Aung et al. [4] has suggested two approaches, namely random forest and random forest-based K-means algorithm. They achieved good accuracy and able to extract the attacked illusion. Abdulhammed et al. [5] had proposed deep and variational auto-encoder (VAE), voting, random forest and stacking machine learning classifiers for anomaly-based intrusion detection of imbalanced network traffic. Ahmad et al. [6] proposed various machine learning techniques such as ELM, SVM and RF for IDS. The analysis shows that the ELM approach got good accuracy, precision and recall compare to other models. ELM is a suitable technique for IDS that is designed to analyze a large amount of data.

Aburomman et al. [7], had analyzed various ensemble and hybrid techniques in the intrusion detection system. Choudhury et al. [8] analyzed various classification algorithms. The analysis shows that RF and BayesNet are most precise when compare to the other algorithms. Chang et al. [9] presented the network IDS using random forest and SVM approaches to improve the accuracy in computer networks. The authors build two machine learning algorithms which are used for improve high detection rate in network intrusion detection.

Chabathula et al. [10] have analyzed IDS system that has using SVM, KNN, J48, random forest, adaboost, nearest neighbors generalized exemplars algorithm, voting features interval classification algorithm and Naivebayes probabilistic classifier. The analysis shows that principal component analysis (PCA) is the most precise among other approaches. Jayveer Singh et al. [11] surveyed different machine learning techniques and soft computing techniques for IDS.

Joshi et al. [12] proposed various classifications, clustering in intrusion detection system. Khan et al. [13] analyzed various classification techniques on intrusion detection systems. Li et al. [14] used hybrid methods such as particle swarm optimization (PSO) and random forest (RF) approaches are used to improve the better performance in detecting the attacks on the network.

Latah et al. [15] built software-defined networking (SDN) controller to improve the accuracy in anomaly-based intrusion detection. SDN is compared with many supervised machine learning approaches. Finally, the analysis shows that the decision tree (DT) is most precise when compared to other models using SDN controller.

Malik et al. [16] proposed a hybrid algorithm named PSO and random forests algorithm. That is, PSO is used for

dimension reduction and RF is used for classification. Murugan et al. [17] analyzed various detection spams in social networks using machine learning such as SVM, NaïveBayes, random forest and decision tree (J48) algorithms.

Maniriho et al. [18] presented a combined machine learning approach with a two-feature selection techniques such as correlation ranking filter and gain ratio feature evaluator for the anomaly NIDS. Tsai et al. [19] reviewed the recent studies of intrusion detection by machine learning techniques. The authors are reviewed a large number of machine learning techniques which are used in the intrusion detection domain review has included single classifiers, hybrid classifiers and ensemble classifiers.

Wang et al. [20] have presented C-ELM approach to network intrusion detection. The authors built several models with fast learning speed in hidden neurons with binary search. Zhang et al. [21] proposed a network intrusion detection system based on data mining algorithm called random forest in anomaly-based, misuse-based and hybrid-network-based intrusion detection systems. Yin et al. [22] proposed a model based on recurrent neural networks (RNN) to improve the accuracy in NIDS.

Ingre et al. [23] have presented different methods of BFGS quasi-Newton backpropagation algorithm and Levenberg–Marquardt (LM) in ANN. ANN is used for supervised classification learning to improve the accuracy and detection rate. Murugan et al. [24] proposed a hybrid algorithm named feature extraction combination of logistic regression and principal component analysis methods to increase the classification accuracy using machine learning algorithms on twitter data.

3 Stacked ensemble learning

Ensemble learning has three types namely, bagging, boosting and stacking. Bagging and boosting are the alternatives of the voting methods. The bagging approach, homogeneous models are taken to predict the class of test data. Initially, homogeneous selected models predictions are recorded. Finally, the class which is predicted by the maximum number of models is assigned to the test data. Similarly in boosting, the models are trained heavily for the misclassified data in the training phase. Finally, the model which is showing maximum accuracy is considered as the classifier for the test data. The stacking is an effective approach because it is a generic framework, which combines many ensemble methods. It has two levels of learning, base learning and meta-learning. In base learning, the initial (base) learners are trained with training data set. After training, the base learners create a new data set for the meta-learner. Then, the meta-learner is trained with new training data set. The trained meta-learner is used to classify the test set. A crucial part in stacking is the selection of a best base learner. That is instead of selecting a single base learner, select many base learners for the training data set. The main difference between stacked and other methods of ensemble techniques is that in stacking meta-level learning-based classification is applied as final classification. Ensemble classification techniques are shown in Fig. 2. This paper has used stacked ensemble model, in which random forest and gradient boost are base learners. The above algorithm summarizes the proposed method for detecting the attacks. Gradient boost is a recent approach, it is an improved version of adaboost. Adaboost is an additive predictive model in machine learning algorithms. Its prediction based on step–by-step forward stage-wise manner that is in initial stage starts with weak learners for the predictions, for each iteration converts the walk learners to strong by increasing higher data weight points. The slighter variation of adaboost is gradianboost, which introduce a new learner at every iteration on existing weak learner rather than increasing data points. The gradient boost approach can boost any differentiable loss function. The regression trees are built at every iteration and individual trees are added sequentially. That is, the next tree is constructed based on the variation between actual value and predicted values. The random boost works on selecting features and target values which can build set of rules to generate multiple decision trees, finally taking the mean of the results. That is, creating the forest with many trees. Build the tree iteratively from remaining features with n nodes. Finally, calculate the average constructed trees.

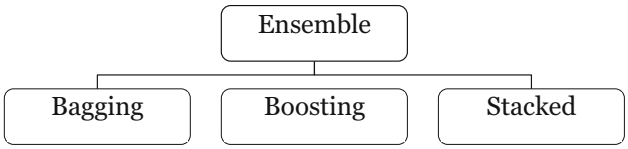


Fig. 2 Ensemble classification techniques

The random forest works on selecting features and target values, which can build set of rules to generate multiple decision trees, finally taking the mean of the results.

Algorithm Stacked Ensemble

1. Input: A training set $D: = (a_1, b_1), (a_2, b_2) \ldots (a_n, b_n)$
   Feature set F: $\{f_1, f_2, f_3, \ldots, f_n\}$
2. Step 1: Learn level-1 classifiers
3. Number of level-1 learners=2
4. Step 2: Learn gradient boosting *(D, F)*
5. Assign *nTrees* = 100

$$f_0(a) = \arg\min_{\gamma} \sum_{i=1}^{m} (\gamma - b_i) = \arg\min_{\gamma} \sum_{i=1}^{m} (\gamma - b_i)^2$$

6. Update the model based on m number of target values
7. **for** $i = 1$ to m do
   $f_i(a) = f_{(i-1)}(a) + g_{(i-1)}(a)$
8. calculate residue $g_i(a)$
9. $g_i(a)$ = predicted value - actual value
10. **end** for
11. Step 3: Learn Random Forest *(D, F, nTree)*
12. G=∅
13. **for** $i = 1$ to nTrees do
14. $T^{(i)} = D_i$
    for i =1 to $T^{(i)}$ do
    t=subset of *F*
    g_i = best feature in *t*
    G ←GUgi
15. **end for**
16. Step 4: construct new dataset of predictions to meta-classifier
17. **for** i=1 to n do
18. $M_h = (a_1', b_1)$, where $a_1' = \{h_1(a_1'), \ldots h_n(a_1')\}$
19. **End for**
20. Step 5: Learn meta-classifier (E)
21. Learn E based on M_h
22. Classify $a_1' \ldots a_n'$ as attacked or normal

4 NSL-KDD data set description

The proposed stacked ensemble model combines gradient boost and random forest approaches. This model is evaluated with NSL-KDD data set [25], which is the new version of KDD-CUP[99] [26, 27].The NSL-KDD data set contains 41 attributes (features) and five different classes. Different classifications in the NSL-KDD data set are shown in Table 1. Attack types present in the NSL-KDD data set and their categorization, which are shown in Table 2. The classes are normal and four different attacks, namely Dos, Probe, R2L and U2R. In the data set, the basic features are from one to ten columns, from 11 to 22 are content features and the rest are traffic features. The proposed model is implemented with R programming language. Experimental result reveals that proposed method increase the attack detection rates and reduces the training time compare to single decision algorithm. Table 3 shows the confusion matrix, which helps to compute the performance measures such as precision, recall, and accuracy.

Table 1 Various classifications of the NSL-KDD data set

KDDTest+	
Normal	9711
Dos	7458
Probe	2421
R2L	2754
U2L	200
Total	22,544

- Denial of Service (DoS) attack: Infinite decline services of a host. Features are source bytes and percentage of packets with errors.
- Probe attack: Unauthenticated information gathering. Features are duration and source bytes.
- User to Root (U2R) attack: Illegal access to local super user or root. Features are number of file creations and number of shell prompts invoked.
- Root to Local (R2L) attack: Illegal access from a remote machine. Features are duration of connection and service requested.

A Dos attack is one of the cyber-attacks, in which the hacker is intentionally change the normal functions of the target system, so that the system become unavailable to its intention users. The main motive of this attack is to denying the requests which are raised from authenticated users. This DoS attack includes flooding, smruf, ping of death and so on. Flooding is one of the DoS attack broadcast the packets to the target machine, so that the machine becomes very busy in receiving the packets. It creates large number of packets and that system is started to deny any request. Likewise, many DoS attacks create the situation of denial of service to the target machine. The proposed approach detects all those attacks and classified as attack. The proposed method has detected 7443 DoS attacks correctly from the NSL-KDD dataset.

Teardrop is a DoS attack, in which the fragmented packets couldn't reassemble by the target machine.

Table 2 Attack types present in the NSL-KDD data set and their categorization

Attack class	Attack type
DoS	Teardrop,Udpstorm,Worm,Mailbomb, Apache2, Back, Land, Neptune, Pod, Processtable, Smurf
Probe	Saint,Satan, Ipsweep, Mscan, Nmap, Portsweep
R2L	Spy, Ftp_write, Phf, Sendmail, Snmpgetattack, Warezclient, Warezmaster, Xlock, Xsnoop, Guess_Password, Httptunnel, Imap, Multihop, Named, Snmpguess
U2R	Perl, Ps, Rootkit, Sqlattack, Xterm, Buffer_overflow, Loadmodule

Table 3 Confusion matrix

Actual	Predicted attack	Predicted normal
Attack	TP	FN
Normal	FP	TN

Teardrop attack feature is specified in feature_wrong_fragment attribute in the NSL-KDD data set. The proposed approach is effectively identifies and classifies this attack. An Udpstorm attack is a UDP flood attack, in which the remote host sends huge number of UDP packets to the target machine. Mailwomb is a DoS attack, in which the large number emails are sent with garbage values to the target system. Land is a DoS attack, in which the attacker sends highly sends the garbage values messages as TCP-SYN packets to the target system. This attack can be possible only when both the sender and receiver IP port numbers are same. The features of the land attack in the given data set are feature_diff_srv_rate, feature_dst_host_serror_rate and feature_dst_host_diff_srv_rate. Neptune is a DoS attack, it is also called half-opened TCP-SYN attack. The attacker is continually sending large number of connection request to the target system. This Neptune attack is recognized using the feature_count, feature_diff_srv_rate, feature_dst_host_serror_rate, feature_src_bytes and feature_dst_host_diff_srv_rate of the NSL-KDD data set. POD is a DoS attack, where the attacker is sending large-sized packets to the target system in a single ping command. The features for this attack are feature_protocol, feature_type and feature_src_bytes and feature_wrong_fragment. Smurf attack is a DoS attack, in which large-sized ICMP packets with garbage values are broadcasted to the target machine. For this attack, the features used in the NSL-KDD data set are feature_protocol, feature_type, feature_src_bytes and feature_wrong_fragment. Probe is another type of attack in IDS, where the attacker scans the target system to find the weakness of the system for exploitation. The probe attack features are feature_service, feature_logged_in, feature_diff_srv_rate, feature_dst_host_count, feature_dst_host_diff_srv_rate and feature_dst_host_ same_src_ port_ rate.

Remote to user attacks (R2U) is another attack in IDS, in which the attacker access the target system as like local user of that system. Then, the attacker can access the target machine and exploit the whole privileges of that system. The features of the R2U attack are feature_service, feature_logged_in, feature_count, feature_same_srv_rate, feature_dst_host_count, feature_dst_host_srv_count and feature_ dst_host_serror_rate in the NSL-KDD data set. User to root attacks (U2R) is an attack, of IDS in which the attacker enter into the target system as a normal user, understanding the privileges of the system and then exploit the vulnerabilities. The features of this U2R attack are feature_service, feature_src_bytes, feature_dst_bytes, feature_hot, feature_num_compromised, feature_srv_count, feature_dst_host_diff_srv_rate, feature_dst_host_same_src_ port_rate and feature_dst_host_serror_rate.

5 Proposed work

Stacked ensemble learning model is also called as multiple classifier system that uses a set of classifiers as base classifiers to build new training data to classify unknown data. Figure 3 shows the architecture of stacked ensemble method. Where are stacked ensemble learning model first select the different base classifiers say B_1, B_2,…B_n, train them using training data set and creates multiple learners L_1, L_2,…L_n from training process. These learner outputs are combined for create a new data set in the form of $<(y'0\ldots y'm), yj>$ for the second level classifier. Here, $y'0$ is a predicted output of the first base classifier on the input $a1'$ and yj is the actual output on the input $a1'$. The second level classifier is also called as meta-level classifier.

The base learners are homogeneous or heterogeneous ensembles. The learners are under same type is homogeneous otherwise heterogeneous. The meta-classifier takes $<(y'0\ldots y'm), yj>$ as an input and trained on that input set. While training on this input set, the meta-learner identify the errors of base learners and adjust them for optimistic solution. This process is repeated with k times for k-fold cross-validation to minimize the error and optimize the output. After repeating this process the meta-learner becomes a generalization model for any input data. In conventional approach, the selected single classifier may perform poorly, that is the classifier behaves good on training data, but it saw an unseen new data, leads to poor classifier. This problem is eliminated in stacked ensemble

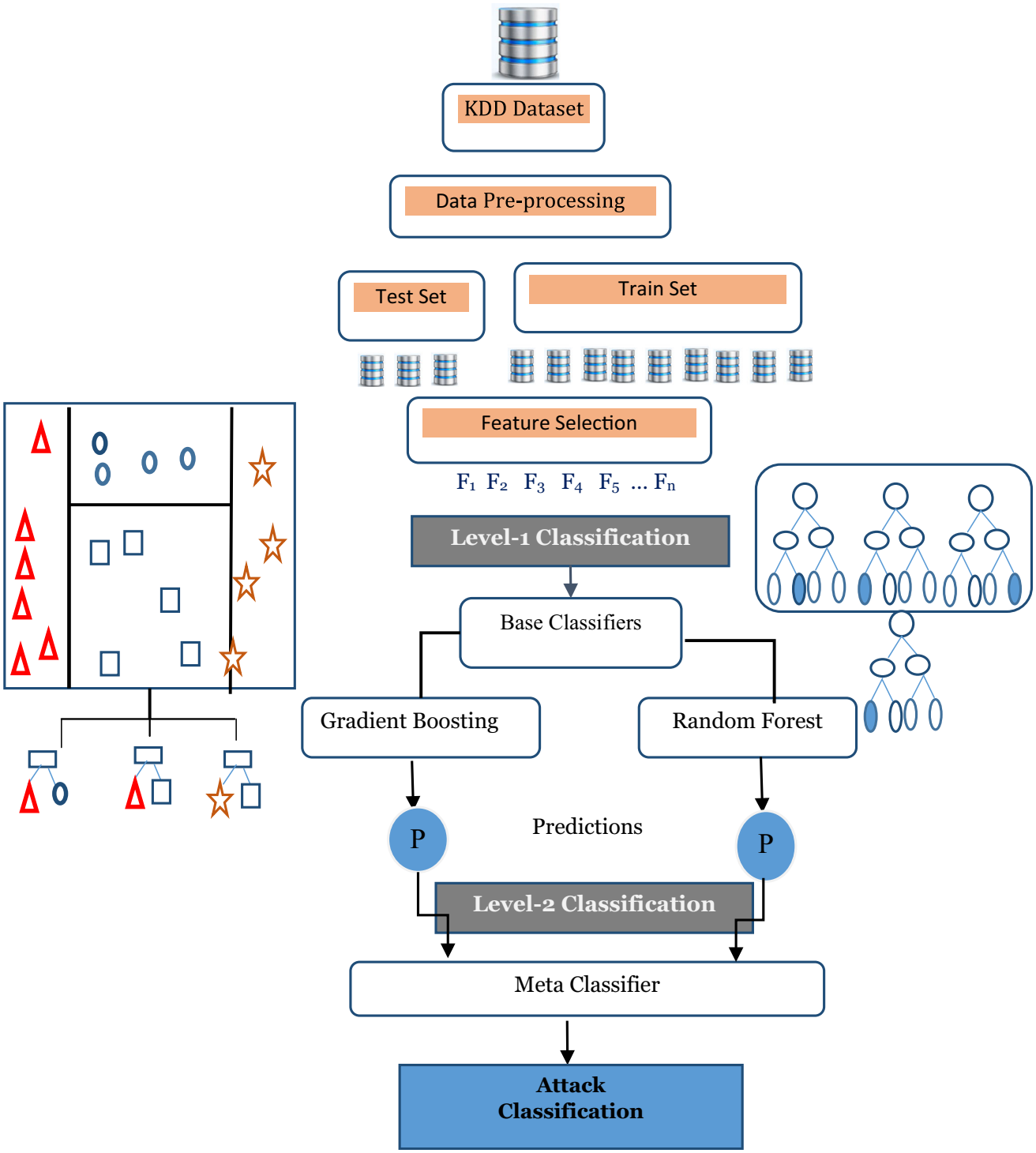


Fig. 3 Stacked ensemble-based IDS architecture

approach. Because even though, the one of selected classifiers in ensemble is unfit for the approach, averaging of all the classifiers can reduce the risk of depending one approach. In general, ensemble method does not guarantee that it can provide best solution for every problem. But it can avoid the risk of poor selection of classifier.

5.1 Performance metrics

In proposed model, the most important performance indicator accuracy. This accuracy is used to measure the performance GBM-RF model in intrusion detection system. The metrics that have been used to evaluate the

performance of the proposed system include the classification accuracy (AC), detection rate (DR), precision and recall. These metrics are expressed by Eqs. 1 to 3 where, true positive (TP), true negative (TN), false positive (FP) and false negative (FN), respectively.

1. True positive (TP)—attack data are classified as an attack.
2. True negative (TN)—normal data are as normal.
3. False positive (FP)—normal data are classified as an attack.
4. False negative (FN)—attack data are classified as normal.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{TN}} \quad (1)$$

$$\text{Detection rate} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{TN}} \quad (3)$$

Table 4 Confusion matrix for the five-category experiments on the KDDTest$^+$

Actual class	Predicted class				
	Dos	Probe	R2L	U2R	Normal
Dos	7443	1	1	0	15
Probe	1438	940	0	0	43
R2L	120	1	2567	2	195
U2R	0	0	7	51	9
Normal	92	36	51	4	9528

Table 5 Results of the performance metrics (detection rate and recall) for the four-category classification of proposed work

	Dos (%)	Probe (%)	R2L (%)	U2R (%)
Detection rate	99.77	38.83	88.98	76.12
Recall	81.85	96.11	97.75	89.47

5.2 Experimental results and discussion

The proposed approach is tested on Intel Core (TM) i5, 8 GB RAM and coding is done by R language. The confused matrix for the five-category experiments on the KDDTest$^+$ is shown in Table 4. Table 5 shows the precision and recall values are obtained by the proposed approach.

Figures 4 and 5 depicts that the proposed method stacked ensemble is obtained higher percentage of detection rate, recall and accuracy compare to conventional machine learning on each attack types. As the gradient boost approach reduces residue error at every iteration, the prediction model improves its performance, thus the attacks detection rate is increased indirectly. The proposed model is compared with several machine learning algorithms and neural network models. The results shows that for each attack type, the RNN and ANN approaches precision and recall values are very less compare to proposed approach.

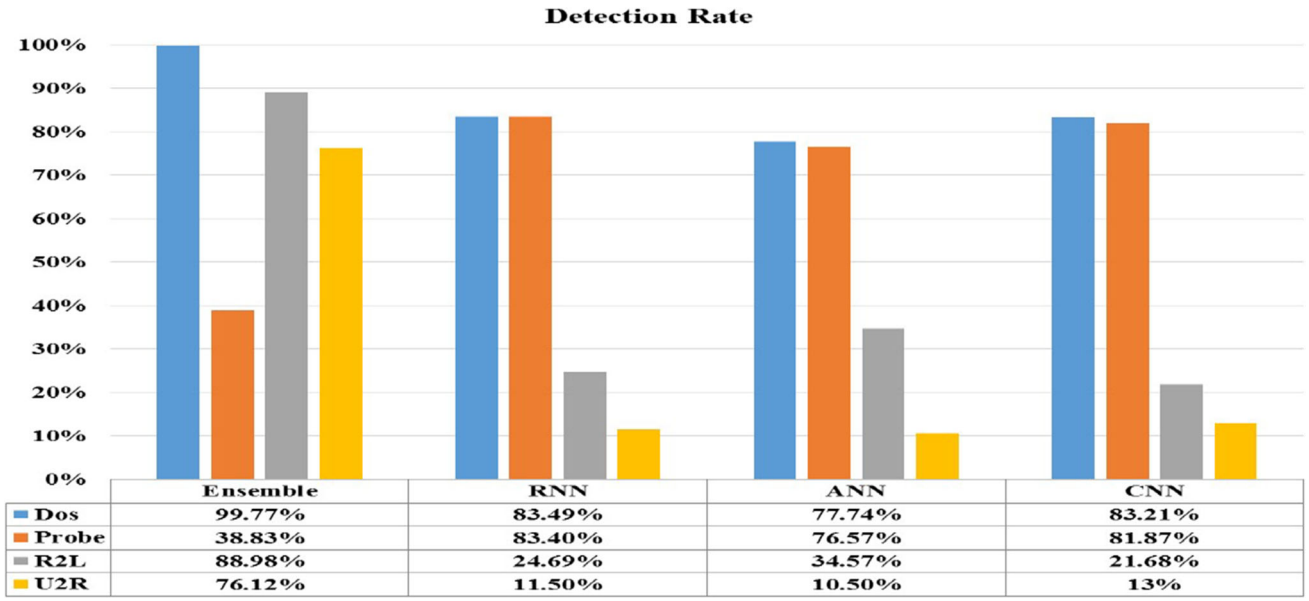
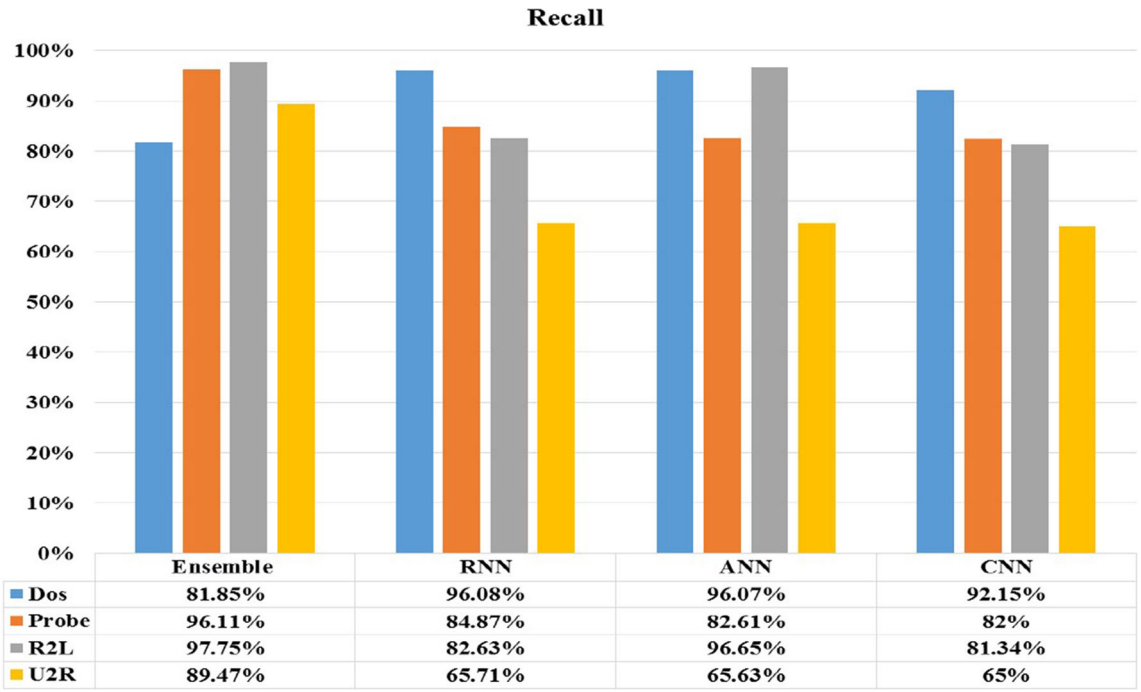
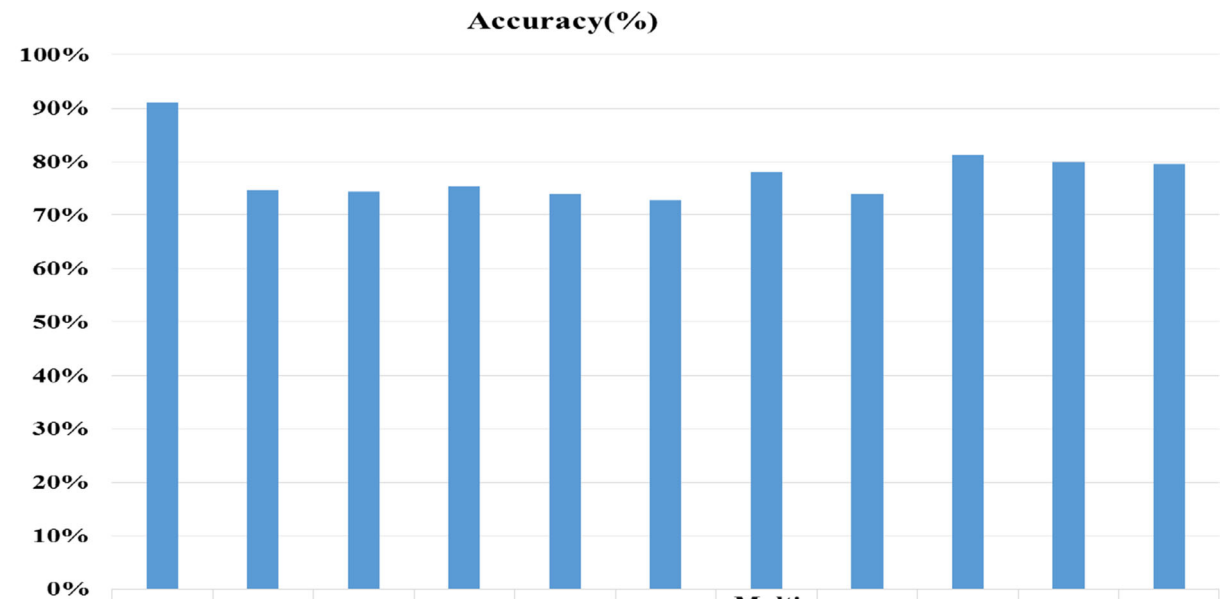


Fig. 4 Different attack detection rate of stacked ensemble with other models



	Ensemble	RNN	ANN	CNN
Dos	81.85%	96.08%	96.07%	92.15%
Probe	96.11%	84.87%	82.61%	82%
R2L	97.75%	82.63%	96.65%	81.34%
U2R	89.47%	65.71%	65.63%	65%

Fig. 5 Different attack recall of ensemble with other models



	Ensemble	J48	Naïve bayes	NB tree	Random forest	Random tree	Multi layer perceptron	SVM	RNN	ANN	CNN
Accuracy(%)	91.06%	74.60%	74.40%	75.40%	74.00%	72.80%	78.10%	74.00%	81.29%	79.94%	79.48%

Fig. 6 Accuracy of proposed stacked ensemble learning and existing approach

In Fig. 6, accuracy is compared with the different machine learning approaches. In the experiment, the machine learning algorithms are used, namely Naïve Bayes, decision tree and random forest, perceptron and neural network models for performance evaluation. The overall accuracy is higher for the stacked ensemble learning method is compared with these above models.

6 Conclusion

The proposed stacked ensemble model is well suited for intrusion detection system to detect the attacks as well as classify them in five different classes. The proposed method is a combined approach of gradient descent and random forest algorithms. The proposed method increases the gradient on the weak learner in every iteration so that the weak learner is converted into a strong enough for predicting test data. The proposed method makes better performance when compare to the other ensemble methods like bagging and boosting. Thus, this paper has deployed stacked ensemble approach to detect the security attacks. The experimental result is also show that compare to machine learning and neural network models, the proposed stacked ensemble model performance is higher. In the future work, stacked ensemble learning classifiers will be used in many machine learning algorithms for the task of intrusion detection method. Furthermore, an investigation of probabilistic, decision tree, non-probabilistic and rule induction-based classification algorithms will be combined as ensemble for the better performance.

Compliance with ethical standards

Conflict of interest There is no conflict of interest from the authors.

References

1. Denning DE (1987) An intrusion-detection model. IEEE Trans Softw Eng 2:222–232
2. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
3. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232
4. Aung YY, Min MM (2017) An analysis of random forest algorithm based network intrusion detection system. In: 2017 18th IEEE/ACIS international conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD). IEEE
5. Abdulhammed R et al (2019) Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic. IEEE Sens Lett 3(1):1–4
6. Ahmad I et al (2018) Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE Access 6:33789–33795
7. Aburomman AA, Reaz MBI (2017) A survey of intrusion detection systems based on ensemble and hybrid classifiers. Comput Secur 65:135–152
8. Choudhury S, Bhowal A (2015) Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection. In: 2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM). IEEE
9. Chang Y, Li W, Yang Z (2017) Network intrusion detection based on random forest and support vector machine. In: 2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC), vol 1. IEEE
10. Chabathula KJ, Jaidhar CD, Ajay Kumara MA (2015) Comparative study of principal component analysis based intrusion detection approach using machine learning algorithms. In: 2015 3rd international conference on signal processing, communication and networking (ICSCN). IEEE
11. Sing J, Nene MJ (2013) A survey on machine learning techniques for intrusion detection systems. Int J Adv Res Comput Commun Eng 2(11):4349–4355
12. Joshi M (2012) Classification, clustering and intrusion detection system. Int J Eng Res Appl (IHERA) 2(2):961–964
13. Khan JA, Jain N (2016) A survey on intrusion detection systems and classification techniques. Int J Sci Res Sci Eng Technol 2(5):202–208
14. Li H, et al (2018) A RF-PSO based hybrid feature selection model in intrusion detection system. In: 2018 IEEE 3rd international conference on data science in cyberspace (DSC). IEEE
15. Latah M, Toker L (2018) Towards an efficient anomaly-based intrusion detection for software-defined networks. IET Netw 7(6):453–459
16. Malik AJ, Shahzad W, Khan FA (2015) Network intrusion detection using hybrid binary PSO and random forests algorithm. Secur Commun Netw 8(16):2646–2660
17. Murugan NS, Devi GU (2018) Detecting spams in social networks using ML algorithms—a review. Int J Environ Waste Manag 21(1):22–36
18. Maniriho P, Ahmad T (2018) Analyzing the performance of machine learning algorithms in anomaly network intrusion detection systems. In: 2018 4th international conference on science and technology (ICST), vol 1. IEEE
19. Tsai CF, Hsu YF, Lin CY, Lin WY (2009) Intrusion detection by machine learning: a review. Expert Syst Appl 36(10):11994–12000
20. Wang C-R et al (2018) Network intrusion detection using equality constrained-optimization-based extreme learning machines. Knowl Based Syst 147:68–80
21. Zhang J, Zulkernine M, Haque A (2008) Random-forests-based network intrusion detection systems. IEEE Trans Syst Man Cybern Part C Appl Rev 38(5):649–659
22. Yin C et al (2017) A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access 5:21954–21961
23. Ingre B, Yadav A (2015) Performance analysis of NSL-KDD dataset using ANN. In: 2015 international conference on signal processing and communication engineering systems. IEEE
24. Murugan NS, Devi GU (2019) Feature extraction using LR-PCA hybridization on twitter data and classification accuracy using machine learning algorithms. Cluster Comput 22:13965–13974
25. https://www.unb.ca/cic/datasets/NSL.html
26. The UCI KDD Archive KDD'99 datasets. Irvine, CA, USA, 1999. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
27. Wu K, Chen Z, Li W (2018) A novel intrusion detection model for a massive network using convolutional neural networks. IEEE Access 6:50850–50859