International Journal of Smart Sensor and Adhoc Network

Volume 3 | Issue 2 Article 8

January 2022

Ensemble Models for Intrusion Detection System Classification

Geethamanikanta Jakka
University of Cumberlands, Williamsburg, gjakka6068@ucumberlands.edu

Izzat M. Alsmadi
Texas A&M University-San Antonio, ialsmadi@tamusa.edu

Recommended Citation

Jakka, Geethamanikanta and Alsmadi, Izzat M. (2022) "Ensemble Models for Intrusion Detection System Classification," *International Journal of Smart Sensor and Adhoc Network*: Vol. 3 : Iss. 2 , Article 8.
DOI: 10.47893/IJSSAN.2022.1209
Available at: https://www.interscience.in/ijssan/vol3/iss2/8

Ensemble Models for Intrusion Detection System Classification

Geethamanikanta Jakka[1], Izzat Alsmadi[2]
[1]University of Cumberlands, Williamsburg, KY 40769
[2]Texas A&M University, San Antonio, TX, 78259
jmani513518@gmail.com, ialsmadi@tamusa.edu

Abstract

Using data analytics in the problem of Intrusion Detection and Prevention Systems (IDS/IPS) is a continuous research problem due to the evolutionary nature of the problem and the changes in major influencing factors. The main challenges in this area are designing rules that can predict malware in unknown territories and dealing with the complexity of the problem and the conflicting requirements regarding high accuracy of detection and high efficiency. In this scope, we evaluated the usage of state-of-the-art ensemble learning models in improving the performance and efficiency of IDS/IPS. We compared our approaches with other existing approaches using popular open-source datasets available in this area.

***Keywords*:** Intrusion Detection Systems, Cyber Security, Data Analytics, Ensemble model

1. Introduction

Nowadays, communication systems play an important role within the life of all the people. Computer networks are effectively used for business data processing, education, and entertainment with internet accessibility and network security growth. All computer systems face security vulnerabilities issues that are technically difficult and economically costly to solve by the manufacturers. Research about finding stacks and removing their effects has been defined as Intrusion Detection Systems (IDS).

"Intrusion Detection Systems" might be helping in detecting the users of the network that aligning several intentions that is by not compromising the security of the network and the host. On the other hand, there are two types of techniques in IDS. "In addition, one is called the Misuse Detection, which is easily understand the overall known attack and works within the online data. Apart from that, the other one is the "Anomaly detection", that can be detected by the abnormal work and behavior within the online data. Analyzing this dataset requires robust tools and techniques to help us find the kind of attack".

- **Working of the Intrusion Detection System:** The system is having the four several steps within the working process of IDS and the system are collecting several data, analysing the overall data, selecting several features, and performing several actions.
- **Collection of several data:** For doing the IDS, there is a requirement in collecting several pieces of information that is about the network traffic such as hosts, different types of traffic and the details of the protocols.
- **Selecting the features:** We need to extract the **necessary** and important features that are also in a larger amount by collecting several data.

- **Analysing the overall data:** The chosen features of the overall data is totally analysed in finding the information that is normal or abnormal.
- **Performing Several actions:** the alarm of IDS or the alert is totally made by the overall system administrator at that time when the total attack is occurred, and it also tells regarding the overall total attack and its types.

The rest of the paper is organized as follows: Literature Review, Dataset description, Overview of Ensemble Model, and it is working, later it defines the features of the data sets.
In addition, it also explains about the several parameters that helps in evaluating the overall performance. Finally, the work's conclusion and future scope are given, and the results and experimental details.

2. Literature Review

A Novel Ensemble Approach for Effective Intrusion Detection System in upcoming areas such as networks, healthcare, and e-governance, machine learning is used to make decisions on criticaldatasets. These algorithms are beneficial and help different fields to make better decisions on datasets. One of the most useful approaches is the ensemble classifier, a combination of classification algorithms. These algorithms are used in machine learning to perform effective classifications. Different classifiers' responses are combined in the ensemble approach to make it a single response [7]. This approach influences achieving better detection than a single classifier's classification accuracy. For the specific dataset, the selection of the ensemble is becoming difficult. A novel ensemble can deal with such a situation, combining the attribute selection algorithm. In addition to this, the novel ensemble is a multiclass support vector machine and KNN classifier. The proposed system performance classification accuracy can be improved by using Incremental Particle Swarm Optimization (IPSO). After optimizing the performance of the proposed system using IPSO, the experiments will be conducted. Five subsets from the standard KDD'99 dataset would be used to conduct the experiments.

"According to [4], the IDS totally monitors the malicious activities or the overall network and gives the forbidden access to all the devices. Apart from that, the IDS are totally used for protecting the integrity and features of several data. On the other hand, the overall proposal was fully used on the NSL-KDD datasets for learning the manner regarding the attacks that are depending on the data mining techniques such as K-means clustering and logistic regression. Therefore, it also generates several rules that are for classifying all the network activities and the overall results shows that the linear regression is effectively detecting the attacks that is about 80%. In addition, the K-means clustering process is showing another kind of results that is with the accuracy around 67%".

"Akash deep Sharma et al. also proposed the intelligent system, and it generally performs the feature ranking that is based on the correlation and gaining the information. Apart from this, the feature reduction process is also made for combining the ranks that are obtained from correlation and information gaining that is by using the novel approach and it helps in identifying the useless and useful features. Hence, the features are then fed for the feed-forward by the neural network that is for testing and training on the dataset of KDD99".

The critical challenge in developing systems for intrusion detection is to adapt it to permit routers and network defense systems to find out if there is any malicious traffic in the network present, especially if it is in the form of network protocols and general access. A novel approach called Spectral Clustering and Deep Neural Network (SCDNN) is proposed in the paper. This approach combinesthe algorithms of spectral clustering (SC) and deep neural networks (DNN) [5]. Asin SC, the cluster centers divide the dataset into k subsets based on sample similarity. Based on the similarity features, the gap between the data point present in the testing region and the training aspect will be evaluated. The evaluated distance will be brought into the deep network algorithms to detect the intrusion. To determine how the model is performing, various critical KDD datasets and a specific sensor dataset have been used in this case study. From the results of the experiment, the SCDNN classifier works better than the backpropagation neural network (BPNN),random forest (RF), support vector machine (SVM), and Bayes tree models. In addition to this, the proposed SCDNN approach provides an effective tool for studying and analyzing intrusion detection in large networks.

A modified stacking ensemble approach to detect network intrusion for years, researchers have faced the issue of detecting intrusions in a network. There are improvements in machine learning that allow researchers for detecting the intrusions regarding the network without using significant databases. The performance of an ensemble method named stacking technique was studied and analyzed on the KDD'99 data set. It is a technique used to combine the different classification models to create a better classifier [2]. The model generation and selection techniques were modified to improve the stacking method. In addition to this, different classifications were used as a combiner method. The subset of the dataset with randomly selected features was used to perform model generations. For the input for the combiner, not all these models were used. In model selection, various metrics were used, and only selected models were used as input for the combiner methodology. When compared to the pure machine learning technique, it is found that the stacking technique used in the experiments brought up better accuracy results all the time. The highest rate of identification for user-to-root attacks was obtained in the experiment, and it is the highest detection rate compared to other studies.

“Ensemble-based Classification Model for Network Anomaly Detection in Massive Datasets Internet traffic” over the network is increasing exponentially. The bulk aspects of Internet traffic gen are the growing concerns of the identification of legitimate users. The normal operations or the functionalities of the network, such as traffic classification to the allocation of resources and managing the services, get affected due to the occurrence of the anomalies. Therefore, it is important to detect the anomalies in the network in each time frame [3]. The choice of specific aspects and the working algorithms used regarding the classification of several patterns of network traffic that impacts the overall efficiency within the detection model. Apart from that, the present solutions do not address problems such as a lack of balance between several classes, dimensionality, and also the variations within the type of anomalies that happen while detecting several anomalies that is within large-scale data. Anomaly Detection in a large data set is proposed to detect the networking datasets of the real-world. It has three models, Hoeffding bound clustering is done for

identifying the optimal subset of features, and eigenvalues computation module to refine the features set, and for the classification of the network traffic the fast decision tree is used. The dataset of KDD'99 is used for the detection rates, the Enclass is 98.58 percent effective, there is a 0.42 percent false-positive rate, and a 96.06 percent F-score.

The approach for Deep Learning is for Network Intrusion Detection within the Software Defined Networking. For the future internet, and Software-Defined Networking is becoming a promising solution. It is cost-effective and helps in being adaptable and dynamic with features that can easily be managed. In addition to this, it is perfect for the high-bandwidth and dynamic nature of present applications. The security of the network can be strengthened using Software Defined Network. It has a logical aspect of the controllers and a global network [8]. On the other side, SDN can also bring tremendous improvement in possible threats. In the SDN environment, anomalies can be detected. This paper has discussed and applied a deep learning idea for detecting anomalies based on the flow. A Deep Neural Network (DNN) model is built for an intrusion detection system. The DDN is applied and used for the NIDS model in the context of the SDN. The DNN model is trained with the NSL-KDD Dataset. Among the forty-one features of the NSL-KDD dataset, only six basic features are used for the work. These basic features of the NDSL-KDD dataset can be easily obtained in an SDN environment. It is confirmed that the deep learning approach shows strong potential and can beused to detect anomalies based on the flow in the SDN environment.

Building the efficient IDS on the that si totally based on the Ensemble Classifier and Feature Selection. In network topology, the Intrusion detection system (IDS) is one of the techniques widely used in the industries. It is used in the protected systems to protect critical assets' integrity and presence [10]. From the machine learning industry, both the supervised and unsupervised learning approaches are used to improve the efficiency of IDSs. Yet, there is an issue with proper intrusion detection aspects in achieving good performance. Many redundant and irrelevant data in the high-dimensional dataset interfere with the classification process. In detecting each type of attack, the individual classifier is less likely to work on it. They are less adaptable for novel attacks built for stale datasets. A new intrusion detection framework is proposed in the paper. The selection of feature and ensemble learning techniques are used in the framework. CFS-BA, a heuristic algorithm, is proposed for dimensionality reduction. Based on the correlation between features, it selects the optimal subset. The random forest (RF), C4.5, and forest by penalizing attributes (Forest PA) algorithms are combined to introduce an ensemble approach. The probability distributions of the base learners for attack recognition are connected using the voting technique. The experimental results reveal that the proposed method exhibits better performance than other approaches in several metrics.

Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing, the adoption of cloud computing is spreading widely. Because of this, cybercriminals are attracted to such services. On the other hand, the Cybercriminals use “Distributed denial of service (DDoS)” attacks to target the bandwidth of the clouds, services, and other aspects. They also render the unavailable clouds for cloud users and for the cloud providers [6]. These types of

attacks are becoming a common form of attack. In cloud DDoS attack defense, the feature choice has been found out as a preprocessing phase. It can increase the accuracy of the classification and reduce the complexity of the computation. During supervised learning, the feature selection identifies the important features of the original dataset. A multi-feature selection method based on ensemble is proposed to combine the output of four filter methods to attain an optimum choice. The benchmark dataset regarding the intrusion detection, NSL-KDD, and also the decision tree classifier was used to perform a proper evaluation based on the experiments of the proposed method. According to the findings, the proposed ensemble-based multi-filter feature selection methodology can effectively lower the number of features from 41 to 13. In addition to this, the proposed method has a better detection and classification precision rate than the other classification techniques.

Active Learning over Evolving Data Streams is using the Paired Ensemble Framework One of the main sources of big data is Stream data. In practice, stream data classification can be challenging because of the inherent scarcity of labeled instances and the underlying concept drift. These challenges in stream data classification can be tackled by proposing a paired ensemble active learning framework. This active learning framework would include three parts. In the first part, an ensemble model with two base classifiers would be used to detect the change over time [9]. It would be used to make predictions on new instances. In the second part, the most informative instances would be found using two active learning strategies. The potential changes that would happen anywhere in the instance space won't be missed. In the third part, a margin-based metric would be used to measure the informativeness of the instances. The uncertain instances would be captured effectively. The labeling cost would be reduced, and the performance of the model won't be compromised by using an active learning paradigm. The experiments were conducted on the datasets of the real world. These experiments demonstrate that the proposed approach can achieve good predictive accuracyon data streams.

According to Pham et al. 2018, an intrusion detection system mainly helps to analyze anomalous behaviors, which are present inside and outside of the network system. To analyze and solve those problems over the IDS system, the application of machine learning has been increased. There are many limitations are present for applying single classifier over the “classification of normal traffic and anomalies”, which helps to find the main idea for structure ensemble or hybrid models for an efficient system [11]. According to the researcher, the main aim of research study is to develop the whole performance rate of the IDS with the help of the fetcher selection and ensemble method. In this paper two techniques such as Boosting and Bagging including a “tree-based algorithm” are used for the ensemble model [12]. In this paper proposed models were evaluated with the help of the NSL-KDD. Researchers present experimental results with this study also.

When dealing with the selection of 35 features extracted, the sweeping prediction method utilizing J48 even as previously known generated satisfactory achievement in terms including both recognition rate and FAR, according to the experimental data.

3. Data Set Description

"The data sets regarding the model are used for calculating the overall model and it can also be used for detecting the attacks that are accurate or not. Apart from that, the quality of the data sets affects the overall outcome of the "Network Intrusion Detection System (NIDS)". For this purpose, we are totally considering that the KDD cup 99 data sets are having the detailed description regarding several features that are given below".

KDD Cup 99 Data Set.

"KDD cup 99 datasets have been excited for more than 15 years. Still, we are using it for academic research. KDD data set is also known as Intrusion Detection Technique. The datasets of KDD cup 99 are totally created by the overall processing of the tcpdump within the overall portion of DARPA 1998 IDS and these datasets are totally created by the "Lincoln Lab under contract to DARP". Apart from that, it was totally intended by simulating the overall traffic seen which was medium sized in the UD Air force base".

"KDD '99 data set is having around 42 attributes and these attributes are used for the experimental study. On the other hand, this data set will me improving within the dataset known as KDD 99 and from here the duplicate was totally removed for getting out of the biased results regarding classification. Hence, this data set is having the total number of 125,973 instances or also the entries".

As shown in Figure 1 KDD data set, out of 42 attributes, 41 are divides in four different groups that are discussed below.

I. "Basics: Features are the attributes of individual TCP connections".

II. "Content: Features are the attributes within a connection suggested by the domain knowledge".

III. "Traffic: Features are the attributes computed using a 2 second time window".

IV. "Host: Features are the attributes designs to assess attacks that last for more than 2 seconds".

Much research is going on the intrusion detection model, where data is used for training and testing. This data set mainly consists of two types of attacks: Normal: 67343, and Anmoaly:58630. The 42 attributes are classified below.

Sr. No	Attribute Name	Sr. No	Attribute Name	Sr. No	Attribute Name	Sr. No	Attribute Name
1	duration	10	hot	23	count	32	dst_host_count
2	protocol_type	11	num_failed_logins	24	serror_rate	33	dst_host_srv_count
3	service	12	logged_in	25	rerror_rate	34	dst_host_same_srv_rate
4	src_bytes	13	num_compromised	26	same_srv_rate	35	dst_host_diff_srv_rate
5	dst_bytes	14	root_shell	27	diff_srv_rate	36	dst_host_same_src_port_rate
6	flag	15	su_attempted	28	srv_count	37	dst_host_srv_diff_host_rate
7	land	16	num_root	29	srv_serror_rate	38	dst_host_serror_rate
8	wrong_fragment	17	num_file_creations	30	srv_rerror_rate	39	dst_host_srv_serror_rate
9	urgent	18	num_shells	31	srv_diff_host_rate	40	dst_host_rerror_rate
		19	num_access_files			41	dst_host_srv_rerror_rate
		20	num_outbound_cmds			42	class
		21	is_hot_login				
		22	is_guest_login				

Fig 1: Kdd99 Data Sets

"KDD Cup'99 is mainly used for building the overall dataset within the "intrusion detection system (IDS)". Apart from that, the dataset of KDD is having two types of critical issues that are totally conducted by the overall statistical analysis and it can profoundly affect the overall performances of the system. Therefore, most of the significant problems within this dataset are many replicated records, about 53% of correctly classified instances and 47% of incorrect classified instances".

Training Methods.

"This dataset contains categorical data of 42 columns, so we will present three methods that work well with categorical data: logistic regression, Naïve Bayes, and decision tree, ensemble model. In the analysis section, we are going to use 80% data for training and 20% data for testing the model. Apart from that, the techniques of statistical techniques, benchmark data set and data mining are used within the overall research work, and it is used for IDS that are totally explained below",

4. Ensemble Model

"More two of the trained models are ensemble together in forming the new and latest models and it is known as the ensemble model. In addition, this model is fully defined as several sets of classifiers that are individually trained and their overall predictions are combined in classifying new and latest data. Furthermore, the ensemble model also combines the main output regarding the classifiers that are produced by the weak learners within the single and composite classification. Hence, the main purpose regarding the combination of all the classifiers is building the overall hybrid model and it will be improving the accuracy of the classification as it is compared with the each of the classifiers".

4.1 Aggregating the bootstrap (Bagging)

"Bagging is the type of the machine learning that ensembles the algorithm that is used for improving several algorithms that are used for regression experiments and statistical classification. On the other hand, it also reduces the variance and also resolves several problems regarding overfitting the overall data. Hence, the overall algorithm is for bagging and the algorithm are as follows",

Algorithm

- "For m = 1 to M//M (iterations)

a. Drawing the sample of bootstrap that is Sm regarding the overall data

b. Learning about the classifier that is to Cm from the Sm

- Example of each of the test sample
    a. Trying all the Classifiers of Cm
    b. Predicting the actual class that totally receives the higher votes"

The dataset D that is given contains the 'n' tuples where the algorithm of bagging creates the 'm', and it is different from the given datasets where each dataset is containing the 'n' tuples by the replacement. On the other hand, this sample is also called the bootstrap sample. Hence, each model is fully fitted with the sample of bootstrap, and it is fully combined by taking out the means of all the outputs.

4.2 Boosting

The actual goal of the boosting algorithm is totally used for the supervised learning, and it is for reducing the overall bias for using the different techniques of classification. On the other hand, the weak learner is the algorithm that is having better accuracy than other random surmises. In addition, the boosting algorithm combines the amount of the weak learners that form the supervised and strong learning algorithm. Hence, the main boosting algorithm is given as",

Algorithm

1. "Initialising the weights that is wi = 1/N (i = 1…. N)"
2. "For m = 1 to M//M…. (No. of iterations)

a. Learning the classifier that is Cm by using the weights of the current examples

b. Computing the estimation of the weighted error

Err = Σwi of all the incorrect ei / Sum of the overall weight"

3. "For each of the example regarding the test:
    a. Trying all the classifiers that is Cm
    b. Predicting the overall class that used to receive the higher sum of the overall weights that is (am)".

"The weak learners are totally stacked within the top of each other and within the new weak of the algorithm is added where the overall data is reweighted".

Accuracy. The Boosting and Bagging are being calculated by using the Accuracy. Where accuracy is the proportion of the total number of corrected predictions cam be computed as fallows [13]:

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

Kappa. The Boosting and Bagging are also being calculated by using the Kappa. Where Kappa is a very good measure that can handle very well both multi-class and imbalanced class problems:

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e}$$

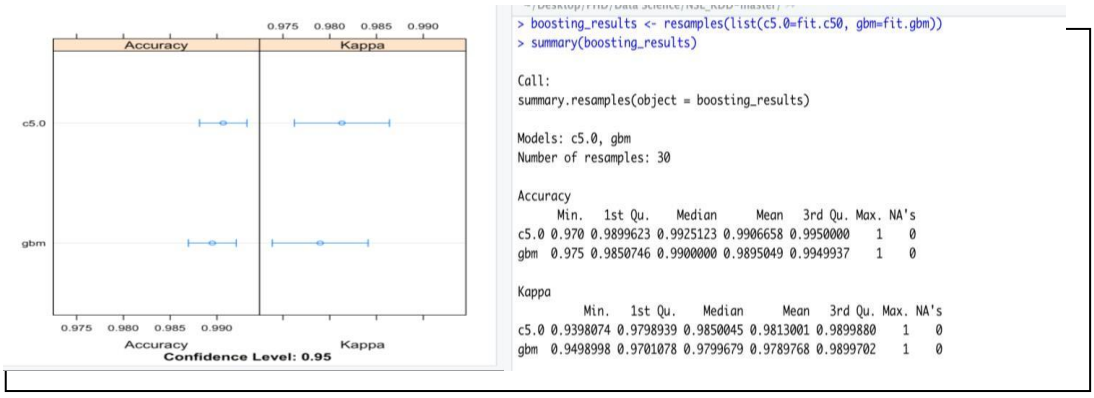

Fig 2: Boosting experiment and result

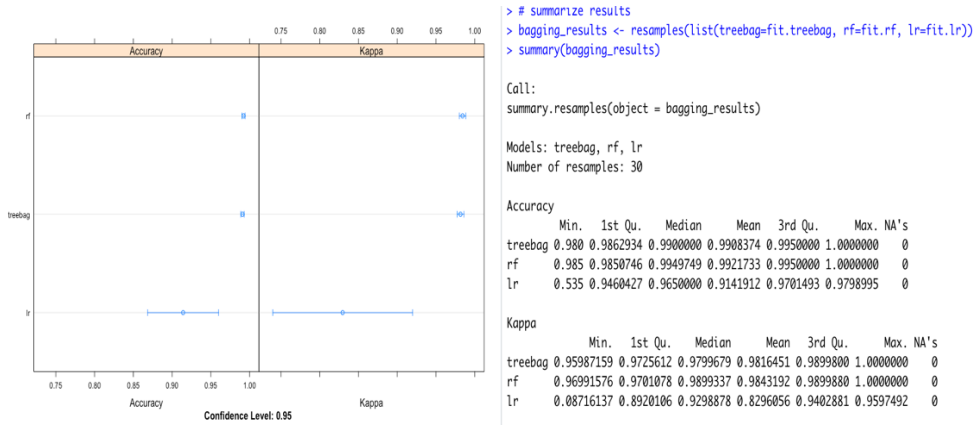

Fig 3: Bagging experiment and results

According to the figure 2 and figure 3, The use of an intrusion detection system or IDS is critical for networking as well as over the device security. IDS are mainly designed to notice traffic patterns and network traffic also to identify attacks. Sometimes this method is called as pattern collaboration. When suspicious activity is located, then an alarm is sent by the system. There are mainly two key processes of intrusion detection such as signature-based intuition method and another one is anomaly based intuition method. The signature-based method compares provided log data and network traffic to known attack patterns to detect potential threats [10]. The anomaly-

based method is intended to identify unexpected threats, such as threats, new malware, and react to them with the help of machine learning.

In this suggested method, classification algorithm was used in data preparation to remove unnecessary features from both the testing and training datasets. After that, a single classification was chosen as the classification algorithm, and the iterative procedure was generated by combining classification model using ensemble procedures, such as Boosting or Bagging. To generate results for assessment, the proposed feature selection model has been tested and trained with resample datasets.

In the figure 2 and 3 it shows about the data accuracy and kappa, where the accuracy level is in the middle of 0.975 to 0.990 and kappa lies in the middle of 0.975 to 0.990. With the help of this idea, it is possible to ensemble more complicated models, which will provide a higher level of accuracy lower rate for the false alarm. Here these figures atre mainly shows about the summarize results, where mean accuracy of treebag is 0.99, median accuracy of tree bag is 0.99.

The overall data is totally divided into the testing and training sets by using the model of simple linear and it will help in making the predictions regarding the yields of training sets that is about the error of root mean squared of 173.3129. On the other hand, if we radially tweak the overall ratios that is for emphasizing the robust model of SVM and we are lowering our errors to around 128.7762.

5. A Comparison Study

The results of our analysis are not fully comparable with the previous research done by other authors. One of the reasons it cannot be compared is the author [1].

The main steps that are followed as the main part regarding the research methodology are followed as”

- “The dataset of KDD is chosen”
- “For simulation process Weka tool is selected here”
- “The Random tree application is used by plotting of H20 train and H20 plotting, split data and test is also used as the main binary classifier which is for the simulation test on the Weka tool, and it classifies the overall instances as the normal or attack”
- “The pre-processing of testing and training of the data files is with around 42 attributes, and it is done for generating only 14 types of new training data files for each of the combinations”.
- Ensemble model

6. Conclusion

“KDD cup 99 Data Set was analyzed using different plotting types and then created a model to predict the results for different types of classes. From this research, we protect crucial information from malicious behaviors or attackers. IDS is an analyzer that can be identified within several information (packets), while it also classifies the overall information that can normally or abnormally. Apart from that, this paper is being proposed for the ensemble model

within IDS. Moreover, this model that is already proposed is having the overall accuracy around 99.49% with the KDD99 data set of 42 features, respectively.

The dataset's size is extensive and requires specific preliminary tasks to massage the overall data before processing it for the analysis. For this research, we used Computer architecture: CPU (Intel I5: Mac OS, Software toolchain: command shell, R programming languages, data mining tools (Weka), data analysis software Supporting software/infrastructure: Libraries and Rpackages. Finally, the main proposed model is the robust classifier within IDS, and it has already achieved the overall highest accuracy". For the future work, we can investigate the overall strategy of ensemble learning in more detail and seek more appropriate ways to measure the informativeness of instances".

References

[1] Chhapolika, P. (2016, July 17). *KDD Cup 99 Dataset*. Amazon. https://rstudio-pubs-static.s3.amazonaws.com/196063_b22a7dcd2ec24413b13b280f4865a2c2.html

[2] DEMİR, N., & DALKILIÇ, G. (2018). Modified stacking ensemble approach to detect network intrusion. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, *26*, 418–433. https://doi.org/10.3906/elk-1702-279

[3] Garg, S., Singh, A., Batra, S., Kumar, N., & Obaidat, M. S. (2017). EnClass: Ensemble-Based Classification Model for Network Anomaly Detection in Massive Datasets. *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. https://doi.org/10.1109/glocom.2017.8255025

[4] Gupta, D., Singhal, S., Malik, S., & Singh, A. (2016). Network intrusion detection system using various data mining techniques. *2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS)*. https://doi.org/10.1109/rains.2016.7764418

[5] Ma, T., Wang, F., Cheng, J., Yu, Y., & Chen, X. (2016). A Hybrid Spectral Clustering and Deep Neural Network Ensemble Algorithm for Intrusion Detection in Sensor Networks. *Sensors*, *16*(10), 1701. https://doi.org/10.3390/s16101701

[6] Osanaiye, O., Cai, H., Choo, K. K. R., Dehghantanha, A., Xu, Z., & Dlodlo, M. (2016). Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. *EURASIP Journal on Wireless Communications and Networking*, *2016*(1). https://doi.org/10.1186/s13638-016-0623-3

[7] Rajasekaran, M., & Ayyasamy, A. (2017). A Novel Ensemble Approach for Effective Intrusion Detection System. *2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM)*. https://doi.org/10.1109/icrtccm.2017.27

[8] Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S. A. R., & Ghogho, M. (2016). Deep learning approach for Network Intrusion Detection in Software Defined Networking. *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*. https://doi.org/10.1109/wincom.2016.7777224

[9] Xu, W., Zhao, F., & Lu, Z. (2016). Active learning over evolving data streams using paired ensemble framework. *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*. https://doi.org/10.1109/icaci.2016.7449823

[10] Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Computer Networks*, *174*, 107247. https://doi.org/10.1016/j.comnet.2020.107247

[11] Pham, N.T., Foo, E., Suriadi, S., Jeffrey, H. and Lahza, H.F.M., 2018, January. Improving performance of intrusion detection system using ensemble methods and feature selection. In Proceedings of the Australasian computer science week multiconference (pp. 1-6).

[12] Illy, P., Kaddoum, G., Moreira, C.M., Kaur, K. and Garg, S., 2019, April. Securing fog-to-things environment using intrusion detection system based on ensemble learning. In 2019 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1-7). IEEE.

[13] Jakka, G. J. K. (2021). Extracting Malware Threat Patterns on a Mobile Platform (Order No. 28776251). Available from ProQuest Dissertations & Theses Global; Publicly Available Content Database. (2624620198). https://www.proquest.com/dissertations-theses/extracting-malware-threat-patterns-on-mobile/docview/2624620198/se-2