International Journal of Engineering Innovations in Advanced Technology
ISSN: 2582-1431 (Online), Volume-6 Issue-2, June 2024

Implementation of Machine Learning Algorithms for Detection of Network Intrusion

Pamidi Prameela
M.Tech Scholor, Department of CSE, GVR&S College of Engineering & Technology,Guntur, India.
Email: pamidi.prameela@gmail.com

Dr Sajeeda Parween
Professor, Department of CSE, GVR&S College of Engineering & Technology,Guntur, India.

***Abstract*- The Intrusion Detection System (IDS) is one of the most well-known and widely used techniques in the modern environment. Its primary functions are to provide a high degree of security, protect very sensitive data, and protect networks from any external threats or attacks. The purpose of intrusion detection systems (IDS) is to protect a specific host or network by analyzing and classifying network traffic as either hostile or legal. Additionally, IDS are designed to recognize any extraordinary actions that may occur. Big Data, cloud computing, and the Internet of Things (IoTs) are examples of some of the new industries that have gained a lot of popularity among consumers as a direct consequence of the rapid growth of the Internet. Because of this, a significant amount of data is produced and distributed on the network on a consistent basis, which may be one of the factors that contributes to an increase in the number of attacks. As a consequence of this, a significant number of academics have chosen to focus their research on intrusion detection systems (IDS) and have made contributions in order to put an end to attacks and threats of this kind. In spite of this, a significant portion of the data that is included within the network traffic may be comprised of qualities or characteristics that are irrelevant to the identification or classification of attacks. Consequently, researchers continue to face challenges when it comes to extracting useful information from this sort of network data and determining whether or not the chosen qualities or traits have the potential to improve the performance of immune system detection systems (IDS). Further, in order to deal with the wide range of threats, intrusion detection systems often need a substantial dataset. The identification of the relevant characteristics is a tough endeavor that must be accomplished in order to improve the accuracy and speed of the intrusion detection system (IDS). In addition, the current intrusion detection system (IDS) makes use of a number of different machine learning (ML), deep learning (DL), and ensemble learning (EL) strategies in order to detect or identify assaults and learn from previous information based on signatures. Despite the fact that these approaches are efficient, they may incur large computational costs due to the fact that they must take into consideration every aspect of traffic while doing so simultaneously. As a result, it is imperative that the issue of including features that are not essential and reducing the cost of computation without sacrificing efficiency be continuously addressed. To begin, FSA was used in this research project to exclude features that were not essential and to zero down on the characteristics that were significant from the NSL-KDD and CICIDS2017 datasets in order to find solutions to these issues. This, in turn, was a contributing factor in the creation of intrusion detection systems (IDS) that are capable of operating in large-scale networks, have cheaper prices, and have higher performances. In order to determine which classifier is the most effective in improving the detection engine of the intrusion detection system (IDS) using NSL-KDD datasets, a variety of classifiers that make use of FSA have been investigated and used.**

***Keywords:* NSL-KDD, Intrusion, Big Data, and Deep Learning**

I. INTRODUCTION

In the society of today, services that are based on the Internet have grown excessively popular over the course of the last several decades. At this point in time, the majority of customers utilize these services from any location and at any time by means of mobile phones, laptops, tablets, and other electronic devices. As a consequence of this, the data that is sent via these networks may include information that is either considered sensitive or essential.

Furthermore, as a consequence of the growth of internet technology, private information is continuously moved between devices and data centers for the purposes of storage and retrieval. As a result of these consequences, the perpetrators of the attacks have a window of time to launch a series of assaults that might put the organization or individual targets in risk. A wide variety of cutting-edge methods are used by attackers in order to take advantage of flaws in system security. It is possible that this may result in illegal access to the system, the exploitation of sensitive or private data, or a breach of client accounts. System administrators and security professionals need to implement cutting-edge security processes in order to safeguard themselves against the threats that are now present. Additionally, as new technologies are developed, such as big data and the Internet of Things, there is a steady increase in the quantity of data or traffic that is being generated.

As a consequence of this, the process of updating the attack signature is rapidly becoming more challenging, time-consuming, and slow as a result of the increasing data traffic on the network. In addition, the task of sorting through such enormous amounts of data in order to find information that is relevant or useful is an essential one for data scientists, commercial firms, and marketers. Research in the field of network security is becoming more popular among academics and scientists as a result of the rising number of people using the internet and the volume of traffic that it generates. Researchers who work in these areas (network security) make an effort to prevent attackers or intrusions from exploiting vulnerabilities in systems or networks in order to get unauthorised access (s). Although over the course of the last

twenty years, a number of preventive technologies have been developed, such as firewalls and antivirus software, in order to safeguard networks and systems from potential assaults such as denial-of-service (DoS), user-to-root (U2R), remote-to-local (R2L), probing, and others. It is for this reason that fundamental security protocols are necessary in order to identify innovative attack types in addition to unwanted traffic or data that has the potential to damage the network or operating system. Among these tools is an intrusion detection system, sometimes known as an IDS [1]. An Intrusion Detection System, often known as an IDS, is a collection of hardware and/or software tools that are meant to gather, analyze, and identify incoming traffic. The purpose of these instruments is to identify unwanted activity in networks and individual systems, as well as masquerade assaults [2] and possible threats. Because of the intrusion detection system (IDS), sensitive data is protected from unwanted access while it is being sent across the network. In order to accomplish these roles and goals, the intrusion detection system (IDS) has to be analyzed, its data needs to be interpreted via the application of certain mathematical or statistical methods, and it needs to alert network administrators and administrators of any suspicious behavior [3].

Intrusion detection systems (IDS) may be broken down into three distinct categories: signature-based, anomaly-based, and specification-based [4]. In spite of the fact that several intrusion detection systems (IDS) have been developed over the course of the last two decades in order to detect and fight against prospective attacks, these systems still do not have sufficient flexibility and scalability, which leaves them vulnerable to covert attacks [5] and a variety of other potential threats. [6] Yahoo stated that it had suffered damages of 350 million dollars as a consequence of two data breaches that happened between the years 2014 and 2016. These breaches affected 500 million user accounts and caused the company to suffer losses. The infections are being subjected to a barrage of sophisticated algorithms that are meant to extract data by snipping it. An intrusion detection system (IDS) zone continues to be dynamic for researchers [7] as the number of attacks (outbreaks) increases along with everything else. In order to evaluate the large volumes of data, extract the relevant features, and identify and categorize the traffic as either normal or assaultive, a powerful intrusion detection system (IDS) is necessary. On the other hand, it could be challenging for intrusion detection systems (IDS) to extract information that is relevant or useful from the enormous amounts of data that are generated by new technologies and sent over networks in order to assess incoming traffic. It is required for intrusion detection systems (IDS) to make use of a big dataset and an FST that is able to eliminate extraneous information and identify the factors that influence attack detection in order to be successful in overcoming these challenges. Furthermore, it is possible that redundant or duplicate objects, as well as noise, might be discovered in enormous databases. Furthermore, taking into consideration a huge dataset may have the unforeseen result of boosting the feature count in direct proportion to the total number of observations. This is a consequence that may not have been intended. There is a possibility that this may lead to a considerable number of false-positive (FP) outcomes. It is possible that not all of the features included in IDS datasets are necessary for detection, despite the fact that many of them give information about anomalies in traffic flow. It is possible that the effectiveness and efficiency of IDS might be improved by choosing a greater number of functional components. There have been a number of studies that have used FSA in this context in the literature in order to increase the efficiency of IDS and overcome the challenges that are associated with data dimension. Filter, wrapper, and hybrid techniques are the three types of FSA, and their classification is determined on the evaluation criteria being used [8]. Which goal functions are known as wrappers? These functions take use of a subset of characteristics by using training models before adding or removing features based on the prior model. Filters, on the other hand, are objective functions that evaluate the information content of data based on the qualities of the data (such as correlation measures and inter-class distance as examples). In conclusion, hybrid strategies incorporate the advantages of prior techniques at various points of the decision-making process so maximizing their effectiveness. Nevertheless, there are a number of circumstances in which it may not always be possible to acquire the ideal subset of features by using filter-based approaches. On the other hand, wrapper-based methods regularly provide great results. As a result of the fact that wrapper-based FSA is more accurate than filter-based FSA and may be tailored for the classifiers, it has been applied in this work [9]. In order to evaluate a subset of attributes, this wrapper-based technique makes use of the classification algorithm. When it comes to accuracy, these approaches often beat filter-based methods. This is due to the fact that they examine attributes and extract the ideal subsets based on input from learning algorithms (such as classifiers) [10]. When utilizing a wrapper-based FSA [11], there are three important factors that need to be taken into consideration: the classifier, the metrics (accuracy, precision, etc.) for evaluating the subset of features, and the techniques for discovering the best feature hybridization. In applications such as intrusion detection systems (IDS), where detection accuracy is considered to be of the utmost importance, wrapper-based feature selection is more effective. The hybrid FSAs that mix filter and wrapper approaches have been created by a number of researchers in an attempt to boost performance while also reducing the complexity of the processing stage. The performance of the hybrid techniques, on the other hand, is not even close to meeting expectations. It is possible that it will be essential to carry out a data cleaning procedure on IDS datasets in order to locate, eliminate, or minimize the number of items that are not necessary. One of the most important uses of machine learning techniques is data mining, which basically involves cleaning up the data. In order to anticipate future attacks, machine learning algorithms may be used to analyze the patterns that have been seen in previous attacks [12]. A powerful intrusion detection system (IDS) may be developed by using machine learning techniques. Clustering and classification are the two essential strategies in machine learning (ML) that are used for the completion of various tasks, including descriptive and predictive tasks.

Classification algorithms anticipate the most likely class, category, for fresh input into predetermined classes.

Clustering, on the other hand, does not pre-define the classes of incoming data during the training phase. In addition, the intrusion detection system (IDS) is responsible for classifying the traffic as either malicious or genuine. This is the reason why classification is always mentioned, but clustering is simply advised to identify the kind of attack. Despite the fact that a number of intrusion detection systems (IDS) that are based on machine learning have been created in recent years, assessing whether a network is normal or abnormal is still a difficult problem to solve. As a result of this, a variety of machine learning strategies have been developed in conjunction with the FSA. These strategies have been used to enhance the intelligence of the IDS and to aid in the resolution of the challenges that were discussed before [13]. Extensive research has been conducted up to this point, and all of it has shown that machine learning-based intrusion detection systems (IDS) perform better in terms of implementation and execution [14]. However, only a small number of models are capable of both achieving high detection rates and low computation costs. This is where the majority of models fall short.

II LITERARURE SURVEY

Regarding the topic of information and data science, dimensionality reduction is a crucial difficulty that has to be addressed. Over the course of the last several decades, a number of IDS models have been developed in order to address the problem of lowering the dimensionality of enormous datasets like KDD99 and NSL KDD for example.

and so forth etc. Some of the research that have been conducted have made use of the FSA in order to address the challenges of data dimensionality and to improve the efficiency of IDS. Despite this, professionals continue to struggle with lowering the dimensionality of data and regulating the amount of time their processing takes [43]. Furthermore, as a result of the substantial increase in network traffic, the scope of possible dangers has also broadened. As a consequence of this, a number of researchers have sometimes used a wide range of machine learning techniques for Intrusion Detection Systems (IDS) that make use of Finite State Automata (FSA).

An investigation of Intrusion Detection Systems (IDS) was carried out by Mukkamala and colleagues [44], who used Support Vector Machines (SVM) and neural networks (NN) in their research. Through the course of the experiment, it was proved that Support Vector Machines (SVM) had a great degree of flexibility and are well suited for efficiently managing large datasets. Learning is a process that requires NN to invest a lot of time. Within the framework of this discussion, Fleuret et al. (2004) used the mutual information approach in order to determine the relevant characteristics. When coupled with a Bayes network, this method's performance is superior than that of SVM. Their study has mostly focused on the overall amount of time required for processing [45]. An investigation on Intrusion Detection Systems (IDS) was carried out by Chebrolu and colleagues in the year 2005. The investigation used technologically sophisticated methods, such as Bayes networks and reverse classification trees. Many different forms of assaults have been effectively identified and detected by them thanks to the collection of twelve essential features from their strategy. Regrettably, the detection rates for U2R assaults do not live up to the expectations that were provided [46]. Chou et al. (2008) used additional feature selection algorithms (FSAs) such as correlation-based feature selection (CFS) and rapid CFS in order to address issues that are connected with high-dimensional data. These issues include redundancy, ambiguity, and uncertainty in the datasets that were acquired. The amalgamation of NB and C4.5 is the means by which the suggested method accomplishes the required qualities. Moreover, inspection of the findings reveals that the recommended fuzzy KNN approach has the potential to boost the rate of detection when compared to other estimators [47]. In addition, Yun and Yang presented an alternative approach to choosing features in the year 2007, after which they made the discovery that decreasing the total number of features might potentially improve the accuracy of learning [48]. Zaman and Karray [49] offered two different ways for selecting qualities or attributes: a Backward Elimination Ranking and a Forward Selection Ranking. Both may be found in the following sentence. A Support Vector Machine, often known as SVM, was subsequently used. The DT C4.5 approach was used by Wang et al. (2009) in order to construct the tree. Following that, they used the Information Gain Ratio (IGR) in order to produce an intrusion rule [50]. In order to investigate the KDDcup99 dataset, Zainal et al. used decision tree (DT) classification strategies in conjunction with filter-based feature selection algorithms (FSAs). These FSAs included relief-F, Chi-square, and information gain (IG). With the use of FST, we were only able to extract twenty, fifteen, ten, and five relevant traits out of the available forty-one characteristics. Based on the data, it was determined that the "IG" technique performed much better than other methods and also improved the functioning of the model [51]. Both the ID3 and NB classifiers, which are examples of hybrid learning methods, were used by Dewan et al. for the IDS. A weight value has been provided to each characteristic in order to facilitate the selection of the features that are relevant. Following that, the maximum depth of the decision tree for each feature was investigated when these weights were taken into consideration [52]. In 2010, Khaing presented a Support Vector Machine (SVM) model that included Recursive Feature Elimination (RFE) and K-Nearest Neighbors (KNN) for the purpose of ranking and choosing features on KDD datasets [53]. The findings reveal that the proposed SVM model is better to the traditional one when the three metrics of accuracy, precision, and false negative rate are taken into consideration. The acronym FNR refers to the False Negative Rate. In this regard, it is important to point out that the F-measure has not been assessed. An strategy that makes use of FSA based on correspondence (CFS) was presented by Nguyen et al. [54], and it was given the name enhanced CFS. It is composed of characteristics that are formed with the help of a variety of limitations and conditions. The KDD 99 dataset was used in order to assess the effectiveness of various estimators and classifiers, including Decision Trees (DT) and Bayesian Networks (Bayes net). In comparison to both methods, the recommended model performed better. In addition, Heba et al. address the challenges of decreasing

processing costs and reducing the amount of features by using Principal Component Analysis (PCA) as a reduction strategy in conjunction with Support Vector Machines (SVM). The experiment indicated that there is a possibility that the effectiveness of IDS might be improved while simultaneously lowering the amount of computing that is required [55]. In addition, Sathya et al. used a statistical method in 2011 in order to investigate the connections that exist between the various characteristics of the KDD 99 dataset. In order to do this, the most essential components had to be determined. Mukherjee et al. introduced the Feature-Vitality Based Reduction method (FVBRM) as a means of determining and selecting relevant features in this particular setting. Following that, NB used these particular characteristics in order to construct the model. In spite of this, the procedure takes a longer amount of time because of the sequential nature of the proposed FST, which involves the deletion of features one at a time until the classifier's accuracy rate reaches a predetermined benchmark number [57]. A Fast Finite State Transducer (FST) was presented by Parsazad et al. in the year 2012. This FST made use of a number of Finite State Automata (FSA), including the correlation coefficient, the least squares regression error, and the maximum information compression index. On the other hand, it has difficulty detecting U2R assaults and can only attain a lesser degree of accuracy, namely 58%. The purpose of the research that Revathi and her colleagues carried out was to determine the effectiveness of various machine learning techniques, including random forests, KNNs, and artificial neural networks (ANN). Following the discovery of fifteen distinct traits, the building of the model included the use of RF, NN, and KNN algorithms. According to the data, radio frequency (RF) method achieves a higher level of accuracy than other alternatives, reaching 98.88%. The accuracy, on the other hand, lowers to 97.94% when RF is employed in conjunction with FST (assuming that all characteristics are excluded) [59].

In the year 2013, Karimi et al. [60] used approaches that included symmetric uncertainty and information growth. The gathering of characteristics for Intrusion Detection Systems (IDS) was maximized by the use of these two approaches throughout research. On the other hand, the results suggest that the detection rate has increased somewhat. The precision of the U2R and R2L, on the other hand, was below minimum standards. According to the findings of a study that was conducted in 2014 by Gao and colleagues, a deep belief network that makes use of a restricted Boltzmann machine performs better than both a support vector machine (SVM) and an artificial neural network (ANN) when given the KDD CUP 1999 dataset. In addition, Kim et al. [62] suggested a hybrid method for identifying assaults using the NSLKDD dataset as the basis for their methodology.

III PROPOSED MODEL IN IDS USING ENSEMBLE LEARNING ALGORITHMS

Figure provides a detailed illustration of the process that will be followed for the framework that is being suggested. There are typically four main processes involved in this approach. These steps are as follows: gathering datasets, initializing and pre-processing data, selecting and reducing characteristics, and lastly, predicting and assessing the results. The majority of these methods are produced by the use of

Sklearn" is a Python package that was developed exclusively for the purpose of machine learning [88], while "Kaggle" is a web platform that is used for the purpose of data analysis and modeling. Users of Windows 11 Pro have the option of employing the Intel(R) Core(TM) i7-10700 CPU 2.90GHz processor in order to upload and analyze data using data analysis models. It is possible for registered users to have a maximum of 16 GB of RAM and 4.9 GB of hard disk drive space.

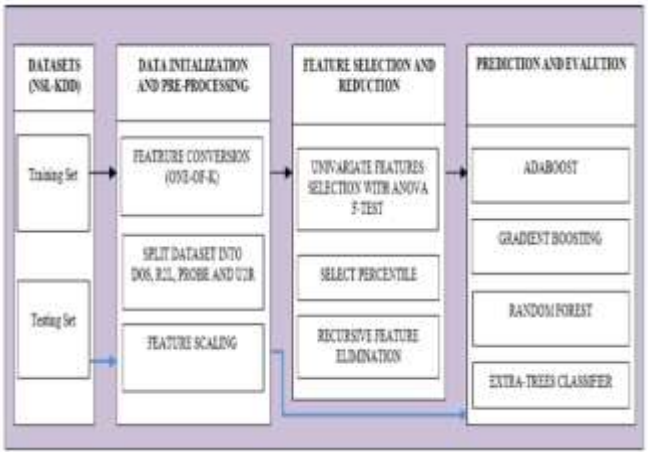

Figure 1: Proposed IDS Frame Work

A. Collection of Dataset

As a result of its creation at MIT's Lincoln Labs [89], the KDD 1999 dataset has been used extensively by academics throughout the course of the last twenty years. There is an improved version of the KDD1999 dataset that is referred to as the NSL-KDD dataset (81). The purpose of this dataset is to address a number of shortcomings.

A dataset referred to as KDD 1999, which may be characterized as follows:

1. Our classifiers are able to deliver objective findings by applying a variety of approaches, which results in enhanced accuracy or detection rate on frequent recordings. This is made possible by the lack of repeated observations in the training or testing sets. The total number of entries (records) in the training and testing datasets are obtained from separate parts of the initial KDD 1999 dataset. This ensures that there is no duplication of information.

There are many factors that support the use of the NSL-KDD dataset, including the following: The capacity of the classifier to deliver objective results is improved when duplicate observations are removed. In addition, both the training dataset and the testing dataset include a substantial quantity of occurrences, which makes it possible to conduct tests on the whole set without the need of picking smaller and smaller pieces at random. In the end, it offers a number of functions, including exact network design, a variety of unfavorable

circumstances, classified observations, thorough packet capture, and a number of other features.

Features Types	Description	Examples
Basic_Features	TCP/IP is the source of the features.	Services, flag, duration land, urgent etc.
Content Features	To gain access to the initial TCP's payload. These features use domain knowledge.	Num_root, Num_shell, hot, Logged_in etc.
Time-based traffic Features	Consider features that span two-second temporal frames have been captured.	Rerror_rate, Srv_count, count, Serror_rate etc.
Host-based traffic Features	These features have the same destination host as the current connections are accessed and span greater than two seconds intervals.	Dist_hst_srv_count, Dist_hst_same_srv_rate etc.

IV **PROPOSED MODEL IN IDS USING MACHINE LEARNING METHODS OVER CICIDS2017 AND NSLKDD**

DATASETS

Because of the complex nature of heavy traffic and the need to strike a balance between a high detection rate and decreased processing charges, it is difficult to develop Intrusion Detection System (IDS) models that are both efficient and cost-effective. As a result, this study proposes a classifier that can be used with FSA.

A framework that is both adaptable and efficient in order to improve IDS detection rates and reduce computing costs. The primary objective of the framework that has been provided is to achieve a high level of accuracy while minimizing computing expenses. The framework that has been suggested consists of five primary steps, as shown in figure 2: datasets, data pre-processing, FSA, model building and assessments, and finally, the analysis and selection stage. Below are detailed descriptions of each of the phases.

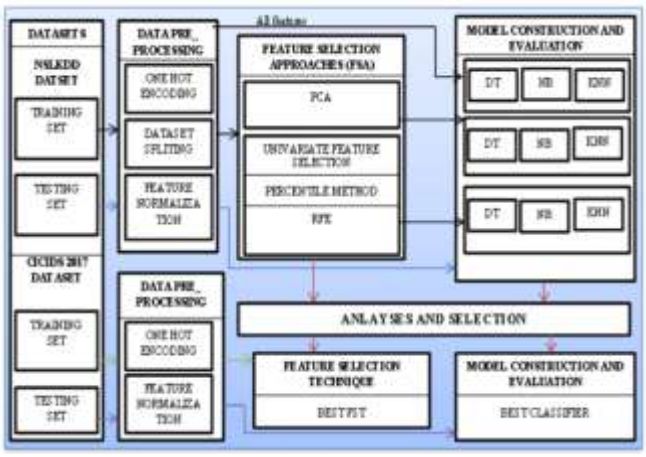

Figure 2: Proposed Framework

In order to conduct an accurate evaluation of the performance of an Intrusion Detection System (IDS), it is required to have a dataset that is consistent in its fundamental characteristics. In addition to this, it is useful for evaluating the differences that exist between the different estimators or classes in IDS. A typical procedure must be completed before beginning the pre-processing activities.

This part provides a comprehensive description of the datasets that are being discussed. These datasets, which are often used for intrusion detection systems, provide a substantial amount of examples that are representative of both regular activities and incursions simultaneously. Detailed descriptions of the NSL-KDD and CICIDS2017 databases are provided in the following paragraphs respectively.

A. CICIDS2017 dataset

A dataset for intrusion detection systems (IDS) known as CICIDS-2017 has been developed by the Canadian Institutes of Cybersecurity. Several incidents that might be classed as sexual assaults are included in the dataset (CICIDS2017), which is comprised of eight files.

Infiltration, Denial of Service (DoS), Brute Force, PortScan, Botnet assaults, Distributed Denial of Service (DDoS), WebAttack, and other sorts of attacks are those that are covered by McAfee, according to a research that was carried out by McAfee [100]. Within this dataset, there are 79 characteristics that serve as representations of the different categories.
The researchers found a great deal of other problems with the dataset, in addition to the fact that it had a significant volume, prominent traits, and the possibility of records being replicated for the purpose of IDS training. In addition, the dataset contains more than eight files that were captured throughout the course of a period of five distinct days. There have also been a number of possible solutions that have been suggested for that specific issue [101]. Additionally, it has a discrepancy that is inherent to it [102]. It is possible that this will lead to estimators that are erroneous and biassed toward the dominant class. The following is a list of some of the issues that are connected to the dataset:

B. Experimental setup and results analyses

The online learning platform known as "Kaggle" and the Python machine learning package known as "Sklearn" were used in this scholarly investigation [88]. On Kaggle, users have the option to upload data-analysis models and then assess such models. The memory capacity of the platform is restricted to a maximum of 16 gigabytes (GB), with exactly 4.9 GB of memory being accessible at the moment.

Putting away. A computer running Windows 11 and a central processing unit (CPU) with a core i5 processor operating at a frequency of 3.6 GHz was used to carry out the whole experiment that was conducted for this inquiry.

Intrusion detection systems, often known as IDS, are very important to the process of network security because of its

capacity to detect assaults at an early detection stage. Nevertheless, the process of gathering and choosing features for identity detection systems (IDS) continues to be both significant and difficult. The effectiveness of the model as well as the computational complexity are both considerably impacted as a result of this. The text provided by the user is simple and accurate.

The fundamental goal of FSA is to completely show or convey a problem or issue by picking a subset of relevant features from the wider category. This is accomplished by selecting attributes from the bigger category.

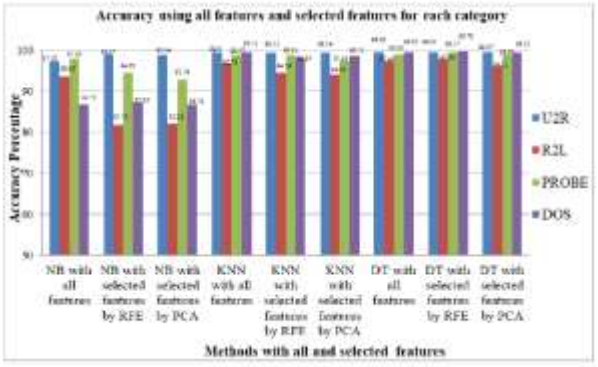

Figure 3 : Various classifiers (KNN, NB, and DT) accuracy utilizing all features and selected features for each class

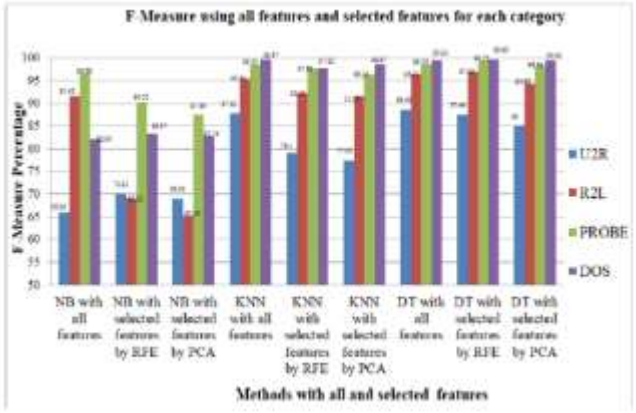

Figure 4: Various classifiers (KNN, NB, and DT) F-measure utilizing all features and selected features for each class

IV CONCLUSION

In order to develop an efficient IDS model, this chapter investigates a wide variety of classifiers that use a variety of FSTs. It has been determined from the findings of the research that decreasing the size of datasets in IDS not only increases the performance of the model but also accomplishes another goal. Lowering the amount of money spent on processing. The results of the NSL-KDD dataset indicate that, with the exception of the U2R attack category, the DT classifier with RFE as FST performs better than other classifiers with FSA in terms of precision, recall, F-measure, and accuracy. This stands in contrast to the performance of other classifiers with FSA. Additionally, the suggested FST has also found a more appropriate and reduced set of features by using ranking methodologies and information gain for the classifiers. This was accomplished by the use of the FST presented. From the results of the research, it was determined that the NSL-KDD dataset had thirteen significant features, whereas the CICIDS 2017 dataset contained eight significant features. In compared to modeling with all of the characteristics, the performance of the model is enhanced, and the amount of computing resources required is reduced. With the use of the Realtime dataset (CICIDS2017), the RFE+DT model was evaluated in terms of recall, G-means, precision, specificity, F-measure, accuracy, training time, and testing time. The recommended model has been compared to other well-known models that have been recorded in the literature in order to offer more proof of the usefulness and productivity of the model that has been suggested. A number of studies have shown that the use of Decision Trees (DT) as the classification method and Recursive Feature Elimination (RFE) as the Feature Selection method (FST) has the potential to improve performance while simultaneously reducing the amount of computer resources required.

REFERENCES

[1]. Zhang, Y., Li, P., & Wang, X. (2019). Intrusion detection for IoT based on improved genetic algorithm and deep belief network. IEEE Access, 7, 31711-31722.

[2]. Elmasry, W., Akbulut, A., & Zaim, A. H. (2020). Comparative evaluation of different classification techniques for masquerade attack detection. International Journal of Information and Computer Security, 13(2), 187-209.

[3]. Shelke, M. P. K., Sontakke, M. S., & Gawande, A. D. (2012). Intrusion detection system for cloud computing. International Journal of Scientific & Technology Research, 1(4), 67-71.

[4]. Rajput, D., & Thakkar, A. (2019). A survey on different network intrusion detection systems and countermeasure. In Emerging Research in Computing,Information, Communication and Applications: ERCICA 2018, Volume 2 (pp.497-506). Springer Singapore.

[5]. Wang, C., Zhao, T., & Liu, Z. (2020). An activity theory model for dynamic evolution of attack graph based on improved least square genetic algorithm. International Journal of Information and Computer Security, 12(4), 397-415.

[6]. Larson, D. (2016). Distributed denial of service attacks–holding back the flood. Network Security, 2016(3), 5-7.

[7]. Vijayakumar, D. S., & Ganapathy, S. (2022). Multistage

ensembled classifier for wireless intrusion detection system. Wireless Personal Communications, 122(1),645-668.

[8]. Alkasassbeh, M. (2017). An empirical evaluation for the intrusion detection
features based on machine learning and feature selection methods. arXiv preprint arXiv:1712.09623.

[9]. Gu, S., Cheng, R., & Jin, Y. (2018). Feature selection for high-dimensional classification using a competitive swarm optimizer. Soft Computing, 22, 811-822.

[10]. Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., ... & Gu, L.(2019). Feature selection based on artificial bee colony and gradient boosting decision tree. Applied Soft Computing, 74, 634-642.

[11]. Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Ala‟M, A. Z., & Mirjalili, S.(2019). Binary grasshopper optimisation algorithm approaches for feature selection problems. Expert Systems with Applications, 117, 267-286.

[12]. Thanh, H., & Lang, T. (2019). An approach to reduce data dimension in building effective network intrusion detection systems. EAI Endorsed Transactions on Context-aware Systems and Applications, 6(18).