Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020)
IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9

Intrusion Detection System using Machine Learning Techniques: A Review

Usman Shuaibu Musa
Department of Computer Science & Engineering
Sharda University
Gr. Noida, UP, India
usmanmusa04@gmail.com

Megha Chhabra
Department of Computer Science & Engineering
Sharda University
Gr. Noida, UP, India
Megha.chhbr@gmail.com

Aniso Ali
Department of Computer Science & Engineering
Sharda University
Gr. Noida, UP, India
Eng.anisafiqi@gmail.com

Mandeep Kaur
Department of Computer Science & Engineering
Sharda University
Gr. Noida, UP, India
mandeep.kaur@sharda.ac.in

***Abstract*—The rapid growth in the use of computer networks results in the issues of maintaining the network availability, integrity, and confidentiality. This necessitates the network administrators to adopt various types of intrusion detection systems (IDS) that help in monitoring the network traffics for unauthorized and malicious activities. Intrusion is the breach of security policy with malicious intent. Therefore, intrusion detection system monitors traffic flowing on a network through computer systems to search for malicious activities and known threats, sending up alerts when it finds those threats. The detection of malicious activities is of two types, the misuse or signature-based detection in which the IDS collects information, analyzes it and then compares it to the attack signatures stored in a large database. The second detection is the anomaly detection which assumes malicious activity as any action that deviates from normal behavior. The proposed paper presents an overview of various works being done on building an efficient IDS using single, hybrid and ensemble machine learning (ML) classifiers, evaluated using seven different datasets. The results obtained by various works were discussed and compared which gives a clear path and guide for future work.**

Keywords— Intrusion Detection System, Machine Learning, security, Anomaly detection, Misuse detection, Classifiers, Ensemble, Hybrid.

I. INTRODUCTION

Intrusion Detection System (IDS) is used to detect threats or malicious activities. The IDS acts as a network level defense to secure a computer network. The intrusion or threat comes in a form of anomaly in a network. Intruders take advantage of network vulnerabilities such as weak security policies, software bugs like buffer overflows, in exploiting the network flaws resulting in the violation of the network's security. The intruders might be system users with less privileges who intend to have more accessing authority or hackers who are common internet users that intend to steal or damage sensitive information from the victim's system [1]. The intrusion detection techniques may be signature based or anomaly detection based. The signature-based detection monitors packet flow in the network and compare them with the previously identified, configured known signatures of known attacks. Whereas anomaly detection technique detects attacks by comparing defined legitimate user parameters with the events that show deviation from the legitimate user parameters [2]. The IDS generates logs and alert the network administrator after the occurrence of malicious activity in a network [24].

The intrusion detection system may be host based IDS (HIDS) or network-based IDS (NIDS). The host-based intrusion detection system are adopted by network administrators to monitor and analyze activities on a particular machine. HIDS often have the advantage over NIDS in the sense that an encrypted information can be accessed when travelling over a network. Its disadvantage is that, HIDS is very difficult to manage as there is need of configuring and managing information for every host. Further, HIDS can be disabled by certain types of denial of service attack. On the other hand, NIDS are software or hardware-based intrusion detection devices intelligently distributed within networks that passively monitors traffic flowing over the network through the devices on which they reside. NIDS have dual interfaces one being used for listening to network conversation and the other for control and reporting. The NIDS have the advantage that monitoring a large network may require a few well fit NIDS and mostly NIDS is invisible to many attackers, thus it is secured against attacks. On the other hand, NIDS have the disadvantage of finding it difficult to detect an attack stroke during a period of high traffic.

The earliest intrusion detection systems developed were majorly signature based, that is, the detection of malicious activities depends on the pre-defined and configured known signatures of known attacks, however this is a major setback as the database of known attack signatures of such intrusion detection systems need to be constantly updated since intruders find a way to exploit network activities on frequent basis [8]. As the machine learning came into existence, it made it possible to perform anomaly detection, which is to detect unknown attacks by comparing legitimate user parameters with events that show deviation from such benign user activities. Over the years several machine learning techniques have been adopted with the purpose of improving the detection rate, reducing false positives and increasing predictive accuracy of IDS. In this research review, it will discover how single, hybrid, and ensemble ML techniques have been performing in the intrusion detection systems [7].

The paper is organized into four sections; the first section covers the introduction of the paper content. The overview of the research papers is discussed in the second section which mentions various ML techniques used for building IDS. The third section provides the comparison of papers based on the result's accuracy, frequently used classification algorithms and datasets. Finally, the discussion, conclusion and considerations for future research work in ML based intrusion detection systems have been discussed in the fourth and fifth section respectively.

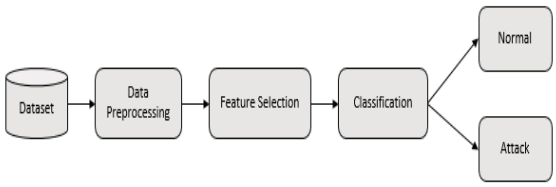

Fig 1: Block Digram of Intrusion Detection System

II. RESEARCH PAPER OVERVIEW

A. Machine Learning (ML)

As a branch of Artificial Intelligence (AI), ML can be defined as a technique in which computers are trained to have the ability to automatically learn and improve or optimize performance criterion without being explicitly programmed, using past experience or example data. Machine learning model focuses on training set of data in accordance to features of interest so that different classes can be predicted [22]. Broadly, machine learning is categorized into supervised, unsupervised and reinforcement learning algorithms.

B. Single Classifiers

Any classifier that is made up of only one classification algorithm is known as Single Machine Learning Classifier. Several intrusion detection systems adopt the use of single machine learning classification models. SVM, Artificial neural network, decision tree, KNN, Naïve Bayes are all single machine learning classifiers and have been applied in different intrusion detection systems studied in this review.

C. Hybrid Classifiers

The hybrid classifier is a combination of two or more ML algorithms for the purpose of boosting the performance of the resulting or aggregated classifier in intrusion detection system. The reason behind employing hybrid approach in the IDS is to amplify its efficiency as it is well proven that hybrid systems perform much more efficient than single machine learning classifying IDS. The first level of hybrid classifier can be represented by either supervised or unsupervised ML algorithms [23].

D. Ensemble Classifier

Ensemble Classifier is a group of multiple machine learning classifiers often called weak learners whose individual decisions are combined in some manner to provide better efficient predictive performance as a consensus decision. Thus, ensemble classifier provides improved performance by aggregating various results of weak learners. Several research works that adopted ensemble techniques show a great degree of accuracy and predictive performance. Methods for constructing ensembles include: bagging and random forest, majority voting, randomness injection, feature-selection ensemble, and error-correcting output coding [10].

III. REVIEW AND COMPARISON OF RELATED WORKS

In the work done by Alkasassbeh and Almseidin [9], three classification techniques were used to address the issues of low accuracy often faced by IDS that adopt artificial neural network with fuzzy clustering when dealing with low frequent attacks. They successfully improved the accuracy by splitting the heterogeneous set of training data into homogeneous subsets of training data thereby reducing complexity of each training set. J48 trees, Multilayer Perceptron (MLP) and Bayes network algorithms were used in the proposed work out of which J48 trees returns the best accuracy. One major drawback of their work is their inability to apply feature selection so as to get rid of all unrelated, redundant and unwanted features.

An intrusion detection system based on single machine learning classifiers was built by Bhavani et al [17] using random forest and decision tree techniques on KDD-NSL dataset. The random classifier returns an accuracy of 95.323%, thus the better result. Low detection as well as false positive rate were not solved by the proposed work [17]. Single Machine learning algorithms were used to detect network intrusion in the proposed work of Ponthapalli et al. The algorithms used in the work are; decision tree, logistic regression, random forest and support vector machine [19]. KDD-NSL dataset was used in the work. The research showed that the intrusion detection system performs best with random forest classifier. They also discovered that the random forest classifier has the least execution time. The proposed work has the limitation of performing efficiently only with a single dataset.

An ensemble-based approach IDS was implemented by Marzia Z. and Chung-Horng L. [10] in which results of multiple supervised and unsupervised machine learning algorithms were aggregated using voting classifier. The work boosts the accuracy and performance of the current Intrusion detection systems. They adopted Kyoto2006+ dataset which is more promising than the most employable KDDCup '99 dataset as it is relatively old. This makes their work to attain a certain level of accuracy but the Recall of the result is quite low in some few cases which indicate high values of false negative rate (FPR).

A real-time hybrid intrusion detection approach was proposed by Dutt I. et al [11] in which misuse approach was used to detect well known attacks while anomaly approach to detect novel attacks. In this work a high detection rate was achieved due to the fact that; patterns of intrusions that were able to escape the misuse detection were able to be identified as attack by the anomaly detection technique. The model's accuracy increased incrementally each day up to a significant value of 92.65% on the last day of the experiment, also, as the model learns and trains the system each day, the rate of false

negative decreases sharply. The issue of slow detection rate persists when the model is applied on a very big size data.

A work done by Verma et al [12] shows that anomaly based intrusion detection has a room for improvement especially in the false positive rate. Extreme gradient boosting (XGBoost) and Adaptive boosting (AdaBoost) learning algorithms were applied on NSL-KDD dataset. Though an accuracy of 84.253 was obtained, an improvement in the performance needs to be done by applying hybrid or ensemble machine learning classifiers.

Some of the previously proposed works have the limitation of inability to apply feature selection on datasets they worked with to eliminate all irrelevant, unwanted and redundant features. In the work proposed by Kazi Abu Taher et al [13], different ML models with different ML algorithms were evaluated with NSL-KDD dataset, feature selection was applied using wrapper method. An improved accuracy was obtained relatively better than the one obtained by the previous works that adopted the same dataset. A major drawback of zero day detection remains unsolved due to the high false positive rate of the model as well as the work focused only on signature based attacks thereby leaving novel attacks undetected.

Some works being done on previous intrusion detection systems lack ability to work efficiently on different datasets. The proposed work of Zhou et al [14] presented a novel intrusion detection system that brings the benefit of combing ensemble classifier with feature selection, this provides an improved efficiency and high accuracy detection of intrusions. The work was carried out using three different datasets; the familiar NSL-KDD dataset and two recently published datasets i.e. CIC-IDS2017 and AWID. For feature selection, CFS-BA based approach was used. The ensemble based approach increases the multiclass classification performance on unbalanced datasets. The model showed the highest accuracy on AWID dataset, giving an accuracy of 99.90%.

Ahmad Iqbal and Shabib Aftab [15] make use of both feed forward neural network and pattern recognition neural network. In addition, they applied Bayesian regularization and scaled conjugate gradient training techniques to train the artificial neural network based IDS. Various performance metrics were used to evaluate efficiency and capacity of the proposed work. The two models were found to outperform each other in different performance measures on various attack detections from the yielded result. Overall, the feed forward artificial neural network provided the better accuracy of 98.0742%. The efficiency of the work needs to be improved by testing the model on different datasets.

An ensemble based approach that combine decision tree, Bayes classifier, RNN-LSTM, random forest was proposed by Vinoth Y. K and Kamatchi K. [16]. This work contributed in handling imbalanced data by choosing the most required effective features to be trained to detect intrusion and send alert to systemadministrators as to whether the intrusion is a normal or abnormal behavior. Though the models performs to some extent of accuracy on NSL-KDD, an experimental trial on the most updated datasets need to be carried out.

Maniriho et al proposed a work on intrusion detection system in which single machine learning classifier (K-Nearest Neighbor) and Ensemble technique (Random committee) were used on two different datasets, NSL-KDD and UNSWN B-15 [18]. A feature selection was applied in this work that generate and used only the most relevant feature subsets for the adopted datasets. The results obtained by the research showed that the ensemble classifier approach performs well over single machine learning technique with a misclassification gap of 1.19% and 1.62% using NSL-KDD and UNSW NB-15 datasets respectively. The issue of large data size, high dimensionality, and standard performance of IDS techniques need to be addressed further in upcoming researches.

A stacking ensemble approach using heterogeneous datasets was proposed by Rajagopal et al. The ensemble technique consists of Logistic regression, K-Nearest neighbor, random forest and support vector machine. The work made use of the most updated dataset in UNSW NB-15 and UGR '16. The UNSW NB-15 was captured in emulated environment while UGR '16 was captured in real network traffic environment [20]. The stacking ensemble approach boosted prediction accuracy and detection speed of the IDS. The model returns the highest accuracy when UGR '16 was used with an accuracy of 98.71%. However, more experiments need to be done on different datasets that include the most recent attack categories.

Perez D. et al, proposed a hybrid network based intrusion detection system (IDS) using multiple hybrid machine learning techniques that work on NSL-KDD dataset [8]. The supervised machine learning technique, Neural Network was combined with unsupervised machine learning, K-Means clustering with feature selection. Another combination was made consisting of support vector machine (SVM) with K-means clustering. The results clearly showed that the combination of such supervised and unsupervised machine learnings complement each other which boosts the performance of IDS. The combination of SVM and K-means with feature selection returns the best accuracy. More hybrid based models need to be built to improve the false positive rate.

A. Comparison Of Related Work

In this research review, several papers have been studied from year 2015 to 2020. Single, hybrid and ensemble classifiers have been widely used in the studied proposed works on the intrusion detection systems. Table 1 describes the comparison majorly in terms of accuracy between different algorithms adopted in the studied research articles.

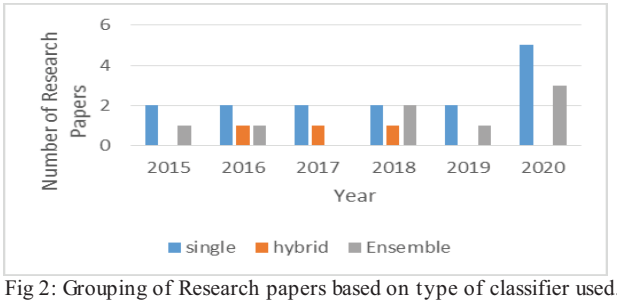

Fig 2: Grouping of Research papers based on type of classifier used.

As it can be clearly observed in figure 3, ensemble classifier returns with highest accuracy whenever it is employed over the years.

B. Datasets Used In The Research Works

Dataset is the collection of instances. An instance is the term used to describe a single row of data. Each instance is made up of multiple features often called attribute of a data instance. The most popular dataset used in the studied work is the KDD-NSL.

In total, seven different datasets have been adopted in those papers that include; KDD Cup '99, KDD-NSL, Kyoto2006+, AWID, CIC-IDS2017, UNSW NB-15 and UGR'16 datasets.

KDDCup '99 data '99 was firstly used for the KDD Cup 99 competition. The dataset has a total of 41 features constituted in each input pattern record which represents TCP connection. The features are both qualitative and quantitative in nature [21]. As the modified version of KDD Cup '99 dataset, the NSL-KDD had solve some of the problems of the original KDD Cup '99 dataset. It is made up of 41 features out of which 38 are numeric attributes while only three are nominal attributes. It also contains 24 training attack types with the dataset having additional 14 attack types. The training set has a total number of 125973 data points. 53.4% of the training set data points are classified as normal connections while the rest (47.6%) are classified as attack [12]. Kyoto2006+ dataset was built based on the real traffic data collected using 348 honeypots in Kyoto University for three years. This dataset contains 24 features out of these, 14 features are the same as in the original KDD Cup '99 dataset. The rest 10 additional features that contains six information related features bring up light to some challenges often faced when KDD Cup '99 dataset is employed [10].

TABLE 2: Distribution of Dataset Usage over the Years

TITLE	ALGORITHM	DATASET	RESULT (ACCURACY)	FINDING	DRAWBACK
IDS using bagging with partial decision tree base classifier[1]	1) Genetic Algorithm (GA) based feature selection. 2) Bagged Classifier with partial decision tree	NLS-KDD99	Bagged Naïve Bayes=89.4882% Naïve Bays=89.6002% PART=99.6991% C4.5=99.6634% Bagged C4.5= 99.7158% Bagged PART=99.7166%	Reduced high false alarm	High time was required to build the model
IDS based on combining cluster centers and nearest neighbors[2]	1) k-Nearest Neighbor (k-NN) 2) Cluster Center and Nearest Neighbor (CANN) 3) Support Vector Machine	KDD-Cup99	CANN=99.76% KNN=93.87% SVM=80.65%	Feature representation was applied for normal connections and attacks	U2L and R2L attacks were not effectively detected by CANN
Comparison of classification techniques applied for network intrusion detection and classification[3]	1) Breadth-Forest Tree (BFTree) 2) Naïve Bayes Decision Tree (NBTree) 3) J48 4) Random Forest Tree (RFT) 5) Multi-Layer Perceptron (MLP) 6) Naïve Bayes	NSL-KDD	BFTree=98.24% NBTree=98.44% J48=97.68% RFT=98.34% MLP=98.53% NB=84.75%	Achieved reduction in false positive	There is need to evaluate the model on the most updated datasets.
Random Forest Modeling for Network IDS[4]	Random forest (RF) based ensemble classifier	NSL-KDD	99.67%	The model is efficient as it returns a low false alarm and high detection rate	A feature selection method like evolutionary computation needs to be applied to improve accuracy
Anomaly Detection Based on Profile Signature in Network using Machine Learning Techniques[5]	1) Genetic Algorithm (GA) 2) Support Vector Machine (SVM). 3) Hybrid Model	KDDCup '99	GA=84.0333% SVM=94.8000% Hybrid (GA+SVM)=98.333%	Low false positive rate	Trials on different datasets need to be done
Fast KNN Classifiers for Network Intrusion Detection System[6]	K-Nearest Neighbor (KNN)	NSL-KDD	99.95%	High accuracy achieved	High computational time due to inability to apply feature selection
Machine Learning Based Network Intrusion Detection[7]	Equality-constrained-optimization-based Extreme learning machines (C-ELMs)	NSL-KDD	98.82%	Improved detection rate and computational speed	The work needs to be carried out on different datasets
Intrusion detection in computer networks using hybrid machine learning techniques[8]	Hybrid model of supervised (Neural Network (NN), Support Vector Machine (SVM)) and unsupervised (K-Means) machine learning algorithms.	NSL-KDD	SVM+K-Means=96.81% NN+K-Means=95.55%	Combination of supervised and unsupervised learning algorithms complement each other in improving IDS performance	Similar approach need to be applied on the most updated datasets
Machine Learning Methods for Network	1) J48 Tress. 2) Multilayer Perceptron (MLP). 3) Bayes Network	KDD '99	J48=93.1083%. MLP=91.9017% Bayes Network=90.7317%	Addressed the issue of accuracy in detecting low	Feature selection was not applied

Intrusions[9]				frequent attacks	
Evaluation of Machine Learning Techniques for Network Intrusion Detection[10]	1) K-Means 2) K-Nearest Neighbor (KNN) 3) Fuzzy C-Means (FCM) 4) Support Vector Machine (SVM) 5) Naïve bayes (NB) 6) Radial Basis function (RBF) 7)Ensemble comprising the six classifiers	Kyoto2006+	RBF=97.54% KNN=97.54% Ensemble=96.72% NB=96.72% SVM=94.26% FCM=83.60% K-Means=83.60%	A more updated and promising dataset in kyoto2006+ was used	Recall of the result is quite low
Real Time Hybrid Intrusion Detection System[11]	Hybrid approach that comprise 1)Frequency Episode Extraction: 2) Chi-Square Analysis	KDD Cup'99	True Positive (TP)=92.65%	The hybrid approach used helped in achieving a high detection rate	The model showed slow detection rate when it was applied on a big size data
Network Intrusion Detection using Clustering and Gradient boosting[12]	1) Extreme Gradient Boosting (XGBoost) 2) Adaptive Boosting (AdaBoost)	NSL-KDD	XGBoost with Clustering=84.253% XGBoost without Clustering=80.238 AdaBoost with Clustering=82.011% AdaBoost without Clustering=80.731%	The work showed that anomaly detection has a room in improving its false positive	The ensemble of the classifiers used needs to be evaluated on the most updated datasets that contains recent attacks
Network Intrusion Detection using Supervised Machine Learning Technique with feature selection[13]	1) Artificial Neural Network (ANN) 2) Support Vector Machine (SVM)	NSL-KDD	ANN=94.02%	High accuracy was achieved due to the application of feature selection	Inability of the work to address the issue of zero day attack due to high false positive rate
Building an Efficient Intrusion Detection System[14]	1)Correlation based feature selection (CFS-BA) 2)Ensemble approach that comprise: C4.5, Random Forest (RF) and Forest by Penalizing (Forest PA)	1)NSL-KDD 2)AWID 3)CIC-IDS2017	Ensemble (NSL-KDD)=99.80% Ensemble (AWID)=99.50% Ensemble (CIC-IDS2017)=99.90%	The model was evaluated on three different datasets and returns with an improved efficiency and high detection rate.	False positive was observed in CIC-IDS2017 dataset
A Feed-Forward ANN and Pattern Recognition ANN Model for Network Intrusion Detection[15]	1)Feed forward Neural Network (FFANN) 2) Pattern Recognition Neural Network (PRANN)	NSL-KDD	FFANN=98.0792% PRANN=96.6225%	The work showed that combining multiple classifiers complement each other in improving performance	The model needs to be evaluated on different datasets to improve its efficiency
Anomaly Based Network Intrusion Detection using Ensemble Machine Learning Technique[16]	1)Decision Tree 2) Bayes Classifier 3) RNN-LSTM 4)Random Forest 5) Ensemble of the 4 classifiers	NSL-KDD	Ensemble=85.20%	The work handled imbalanced data and selected only required features which greatly helped in reducing high false positive rate	Trial on the most updated datasets needs to be carried out
Network Intrusion Detection System using Random Forest and Decision Tree Machine Learning Techniques[17]	1) Random Forest (RF) 2) Decision Tree (DT)	NSL-KDD	RF=95.323% DT=81.868%	Easily implemented	Slow detection rate and high false positive
Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches[18]	1)Single Machine Learning Classifier (K- Nearest Neighbor (KNN)) 2) Ensemble Technique (Random Committee (RC))	1)NSL-KDD 2) UNSW NB-15	NSL-KDD using 1)KNN=98.727% 2) NSL-KDD using RC=99.696% 3) UNSW NB-15 using KNN=97.3346% 4) UNSW NB-15 using RC=98.955%	Ensemble approach generate better accuracy than single classifiers. The model was evaluated using two different datasets.	Fail to address the problem of data high dimensionality
Implementation of Machine Learning Algorithms for Detection of Network Intrusion[19]	1)Decision Tree (DT) 2) Logistic Regression (LR) 3) Random Forest (RF) 4) Support Vector Machine (SVM)	KDD-NSL	1) RF=73.784% 2) DT=72=303% 3) SVM=71.779% 4) LR=68.674%	Showed that working with random forest in building IDS saves execution time	The model performs efficiently only with single classifier

A Stacking Ensemble for NIDS using Heterogeneous Datasets[20]	Stacking Ensemble technique that comprises: KNN, LR, RF and SVM	1)UNSW NB-15 2) UGR '16	1) UNSW NB-15=94.00% 2) UGR '16=98.71%	Boosted prediction accuracy and detection speed was observed	The work needs to be evaluated on multiple datasets

Other datasets include AWID, an acronym stands for Aegean Wi-Fi Intrusion Dataset, consists of real traces of both normal and malicious data obtained from real network environment. AWID was publically available in 2015 as a collection of sets of Wi-Fi network data (Zhou et al, 2019). The CIC-IDS2017 dataset contains normal and the recent common attacks. It was established by Canadian Institute for Cyber security (CIC) in the year 2017. It is one of the newest intrusion detection dataset. It is made up of 2,830,743 records distributed on 8 different files and each record has 78 different labelled features [14]. The UNSW NB-15 dataset was established by a cyber-security research group at the Australian center for cyber security. The acronym UNSW NB-15 stands for University of New South Wales. The dataset has a total of 47 features with two class labels [20].

Table 2 and fig 4 show how the datasets have been adopted by the studied research articles over the years. NSL-KDD has been used a total 14 times which makes 58.33% of the total datasets usage. It was followed by the original KDD Cup 99 dataset which was used 4 times. The UNSW NB-15 was used twice whereas each of Kyoto2006+, AWID, CIC-IDS 2017 and UGR '16 have been used once.

TABLE 2: Distribution of Dataset Usage over the Years

Year	KDD Cup 99	NSL-KDD	CIC-IDS 2017	UNSW NB-15	UGR '16
2015	1	1	0	0	0
2016	1	2	0	0	0
2017	0	3	0	0	0
2018	2	1	1	0	0
2019	0	3	0	1	0
2020	0	4	0	0	2

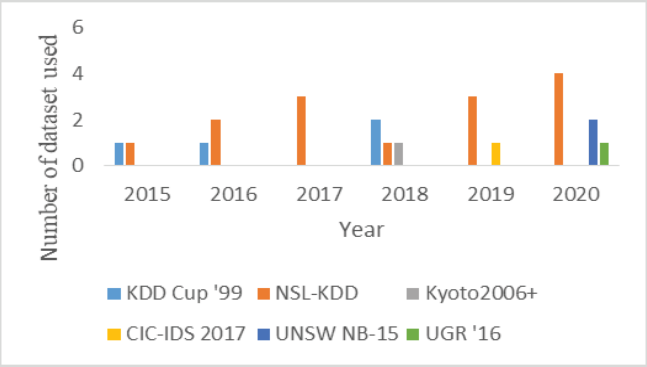

Fig 4: Distribution of dataset usage over the years

IV. DISCUSSION AND FUTURE WORK

As it is shown in fig 3, ensemble and hybrid classifiers have better predictive accuracy and detection rate than single classifiers. For future research work, the following issues have been identified and need to have more consideration in order to improve the performance of intrusion detection systems:

Single machine learning classifiers perform better when they are combined in a specific manner, therefore the hybrid and ensemble machine learning classification techniques need to be used more often.

Some classifiers performs better on specific datasets, in the coming researches, more models need to be developed in such a way that they can be able to perform efficiently on multiple datasets.

A few of the studied research articles have not applied feature selection before the classification stage whereas others have adopted the use of feature selection approaches. The feature selection has to be considered in the coming researches in order to get rid of irrelevant, unwanted and redundant features to improve the efficiency and detection rate of IDS.

As it was discussed in section 3 that 58.33% of the studied research articles had KDD-NSL as dataset. More recently updated datasets need to be worked with in the future research for the purpose of dealing with the most recent malicious intrusions and attacks.

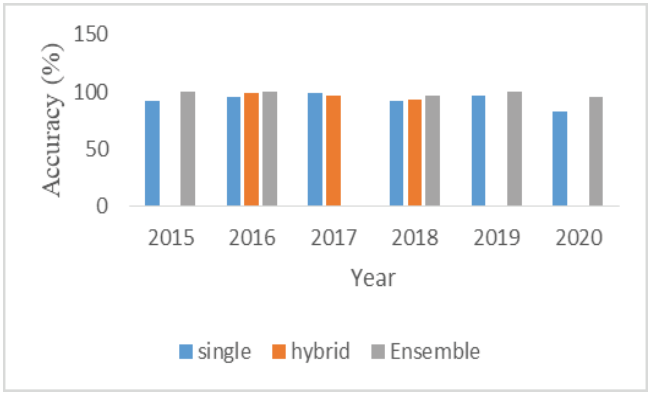

Fig 3: Comparison of Classifiers in terms of Accuracy

V. CONCLUSION

The emergence of machine learning presents new techniques for intrusion detection systems in which various types of classifies have been adopted by researchers and scholars in building intrusion detection systems models. This paper presented various research papers related to using machine learning classifiers in intrusion detection systems published from 2015 to 2020. Among the various models applied in the studied research papers, ensemble and hybrid classifiers have been able to surpass their single classifier counterpart and hence have the better predictive accuracy and detection rate.

REFERENCES

[1] D.P. Gaikwad and Ravindra C. Thool. (2015). Intrusion detection system using bagging with partial decision tree base classifier. Procedia Computer Science 49 (pp. 92-98). Elsevier.)

[2] W. –C. Lin, Shih-Wen K. Chih-Fong T. (2015). Intrusion detection system based on combining cluster centers and nearest neighbors. Knowledge-Based Systems 78 (pp. 13-21). Elsevier.

[3] A.S.A. Aziz. (2016). Comparison of classification techniques applied for network intrusion detection and classification. Journal of Applied Logic 24. Elsevier, 109-118.

[4] Nabila Farnaaz and M.A Jabbar. (2016). Random Forest Modeling for Network Intrusion Detection System. International Multi-conference on information processing (IMCIP) 12 (pp. 213-217). Elsevier.

[5] Kayvan A. Saadiah Y. Amirali R. and Hazyanti S. (2016). Anomaly Detection Based on Profile Signature in Network using Machine Learning Techniques. IEEE TENSYMP. (pp. 71-76). IEEE.

[6] Bobba Brao and Kailasam Swathi. (2017). Fast KNN Classifiers for Network Intrusion Detection System. Indian Journal of Science and Technology. 10(14). Researchgate. (1-10).

[7] Chie-Hong L. Yann-Yean S. Yu-Chun Lin and Shie-Jue L. (2017). Machine Learning Based Network Intrusion Detection. 2nd IEEE International Conference on Computational Intelligence and Applications. (pp. 79-83). IEEE.

[8] Deyban P. Miguel A. A, David P. A, and Eugenio S. (2017). Intrusion detection in computer networks using hybrid machine learning techniques. XLIII Latin American Computer Conference (CLEI). (pp. 1-10). IEEE

[9] Alkasassbeh and Almseidin. (2018). Machine Learning Methods for Network Intrusions. International Confrernce on Computing, Communication (ICCCNT). Arxiv

[10] Marzia Z. and Chung-Horng L.(2018). Evaluation of Machine Learning Techniques for Network Intrusion Detection. IEEE. (pp. 1-5)

[11] Dutt I. et al. (2018). Real Time Hybrid Intrusion Detection System. International Conference on Communication, Devices and Networking (ICCDN). (pp. 885-894). Springer.

[12] Verma P, Shadab K, Shayan A. and Sunil B. (2018). Network Intrusion Detection using Clustering and Gradient Boosting. International Conference on Computing, Communication and Networking Technologies (ICCCNT). (pp. 1-7). IEEE.

[13] Kazi A., Billal M. and Mahbubur R. (2019). Network Intrusion Detection using Supervised Machine Learning Technique with feature selection. International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). (pp. 643-646). IEEE.

[14] Yuyang Z., Guang C., Shanqing J. and Mian D. (2019). Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier. Computer Networks. Doi: https://doi.org/10.1016/j.comnet.2020.107247

[15]] Iqbal and Aftab. (2019). A Feed-Forward ANN and Pattern Recognition ANN Model for Network Intrusion Detection. International Journal of Computer Network and Information Security, 4. Researchgate (19-25)

[16] Vinoth Y. and Kamatchi K. (2020). Anomaly Based Network Intrusion Detection using Ensemble Machine Learning Technique. International Journal of Research in Engineering, Science and Management. IJRESM. (290-296).

[17] Bhavani T. T, Kameswara M. R and Manohar A. R. (2020). Network Intrusion Detection System using Random Forest and Decision Tree Machine Learning Techniques. International Conference on Sustainable Technologies for Computational Intelligence (ICSTCI). (pp. 637-643). Springer.

[18] Maniriho et al. (2020). Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches. International Journal of Intelligent Engineering and Systems. INASS. (433-445)

[19] Ponthapalli R. et al. (2020). Implementation of Machine Learning Algorithms for Detection of Network Intrusion. International Journal of Computer Science Trends and Technology (IJCST). (163-169).

[20] Rajagopal S., Poornima P. K. and Katiganere S. H. (2020). A Stacking Ensemble for Network Intrusion Detection using Heterogeneous Datasets. Journal of Security and Communication Networks. Hindawi. (1-9).

[21] Bolon –C.V. (2012) Feature Selection and Classification in Multiple class datasets-An application to KDD Cup 99 dataset. https://doi.org/10.1016/j.eswa.2010.11.028

[22]] Farah N. H et al. (2015). Application of Machine Learning Approaches in Intrusions Detection Systems. International Journal of Advanced Research in Artificial Intelligence. IJARAI. (9-18).

[23] N. F. Haq et al. (2015). An Ensemble framework for anomaly detection using hybridized feature selection approach (HFSA). Intelligent System Conference. (pp. 989-995). IEEE.

[24] S. Thapa and A.D Mailewa (2020). The Role of Intrusion Detection/Prevension Systems in Modern Computer Networks: A Review. Conference: Midwest Instruction and Computing Symposium (MICS). Wisconsin, USA. Volume: 53. (pp. 1-14).