www.mecsj.com/

Multi-Knowledge Electronic Comprehensive Journal For Education
And Science Publications (MECSJ)
Issue (41),2021
ISSN: 2616-9185

A comparative study for Intrusion Detection Methods Using Machine Learning

Reem A KH AlMeshal
The Public Authority for Applied Education and Training (PAAET)
Email: reem_almeshal@hotmail.com

Abstract

Efficient intrusion detection is needed as a defense of the network system to detect the attacks over the network. Intrusion detection is so much popular since the last two decades, where intruders attempted to break into or misuse the system. There are many techniques used in IDS for protecting computers and networks from network-based and host-based attacks. This paper presents a comparative study for intrusion detection (IDs) using the machine learning (ML) methods, we spotlight the methods used and the results achieved.

Keywords : Intrusion Detection (IDS), Machine Learning (ML), Deep Learning.

الملخص

في الآونة الاخيرة، أصبحت الحاجة ملحة إلى إيجاد اساليب وطرق حديثة لاكتشاف التسلل والهجمات عبر الشبكات كنوع من الدفاع. لقد أصبح اكتشاف التسلل امرا شائعًا للغاية في العقدين الماضيين، حيث يحاول المتسللون اقتحام الأنظمة أو إساءة استخدامها كنوع من التطفل. هناك العديد من التقنيات المستخدمة في أنظمة كشف التسلل (IDS) لحماية أجهزة الكمبيوتر والشبكات من الهجمات القائمة على الشبكة والهجمات القائمة على المضيف. تقدم هذه الورقة دراسة لمقارنة الطرق المختلفة من اكتشاف التسلل (IDs) باستخدام أساليب التعلم الآلي (ML) ، ونلقي الضوء على الأساليب المستخدمة والنتائج التي تم تحقيقها.

الكلمات المفتاحية: كشف التسلل، التعلم الآلي ، التعلم العميق.

1. Introduction

Over the last few years, networks have played a big role in the modern style of life. Cybersecurity has therefore become a fertile area of researches that created a new foresight in data innovation (Alomari et al. 2018). Networks fields, especially security; are one of the most important issues in the field of information security, so it is becoming a primary need of modern society to protect private information flowing over the networks (Kumar et al. 2016).

Cybersecurity methods mainly include different software, such as anti-virus, firewalls, and intrusion detection systems (IDSs). However, many challenges still face computer engineers because hackers and intruders can make successful attempts to break into the networks or computer systems. Because of this, the need to create more powerful Intrusion Detection Systems (IDS) is on the rise in this field (Alsawy et al. 2018).

An IDS is known as a process of controlling the events occurring in the different systems and networks and thus analyzing them for marking possible accidents (Kaur et al. 2013). It is required to ensure the security of the network (Biswas, 2018). IDS is functioning as monitoring the activities in a given environment and identifying whether these activities are "malicious ("intrusive") or "legitimate ("normal") given features gained from the network traffic data (Karpagam et al. 2015).

2. Intrusion Detection

Anderson (1980) proposed the first technique considered intrusion detection system. Many products have evolved since then, but they all still suffer from high rates of false alarms and generate many alerts for low non-threatening acts. In this case, the increasing number of false alerts has increased the burden on security analysts to only detect the seriously harmful attacks on the network. As the network environments update quickly, this allows different and novel attacks to arise permanently, thus creating new challenges in network security with existing IDSs that could not detect some unknown attacks (Lang et al. 2019). For this, many investigators in the field have focused their attention on developing IDSs that would ensure higher accuracy of detection rates of true attacks and reduce false alarm rates (Biswas, 2018).

In general, intelligent intrusion systems or attack systems include collecting information, probing, scanning, remote to local access, the user to root access, vulnerabilities, and launch attacks (Kumar et al. 2016). To reduce the risks for these systems, several techniques have been designed to detect the incursion of the networks.

IDSs and IPSs ("Intrusion Prevention Systems") are the most important protection tools against the developed an ever-growing network attacks. They monitor the state of software and hardware’s running in the network (Lang et al. 2019). Due to the lack of trustworthy tests and validation knowledge, existing IDSs still have problems in improving the accuracy of detection, detection of unknown attacks, reduction of false alarm rates. However, anomaly-based intrusion detection methods are still distress from consistent and accurate performance development.

3. Machine Learning with Intrusion Detection

IDS is considered one of the most important research fields in network security (Juma et al. 2016). Many tools such as firewall, intrusion prevention system and IDS have been designed to stop internet-based attacks. ML is one of the artificial intelligence (AI) branches that acquires knowledge from training data based on known facts. ML is defined as a technique that allows computers to gain knowledge automatically without being programmed as aforesaid by Arthur Samuel in 1959 (Haq et al. 2015). It is categorized into three broad categories: "supervised learning, unsupervised learning, and reinforcement learning".

In "supervised learning (classification)" the instances (features) are classified in the training phase. There are several "supervised learning" algorithms, including: "Artificial Neural Network, Bayesian Statistics, Gaussian Process Regression, Lazy learning, Nearest Neighbor algorithm, Support Vector Machine, Hidden Markov Model, and Bayesian Networks" (Akhilesh et al. 2014).

In "unsupervised learning", the data instances are unclassified. A notable way for the learning method using unsupervised learning depends on the clustering technique. Some of the common unsupervised learners are Cluster analysis "K-means clustering, Fuzzy clustering, Hierarchical clustering, Self-organizing map, Apriori algorithm, Eclat algorithm, and Outlier detection" (Local outlier factor).

In "reinforcement" learning, the computer interacts with an environment to realize a confirmed goal. The ML algorithms used in IDSs are shown in Figure (1) below:

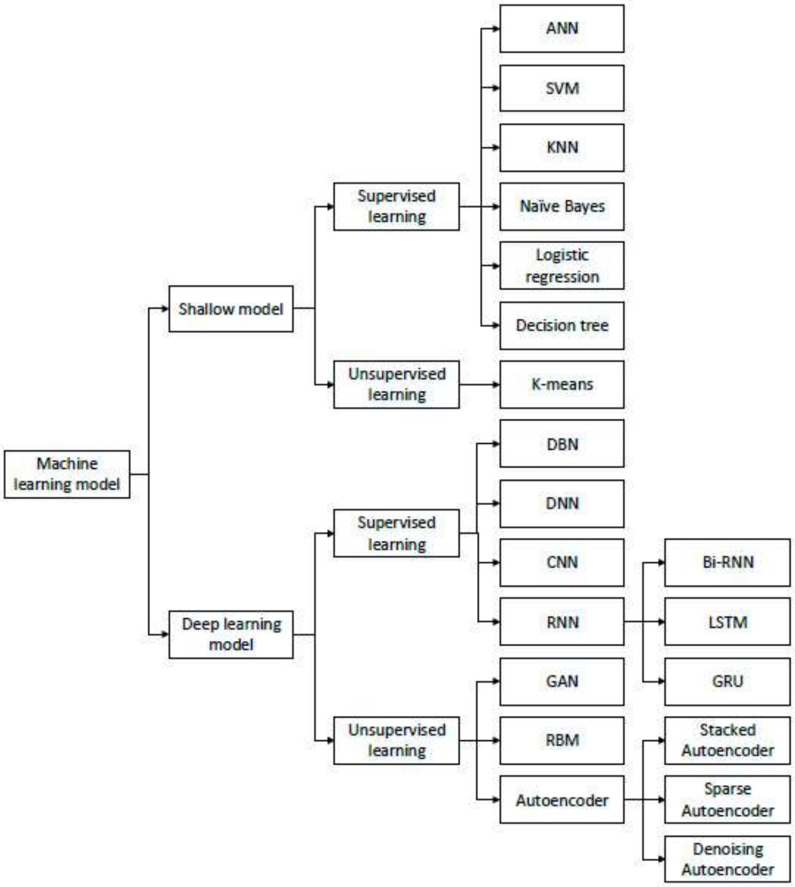

Figure 1: The popular ML techniques (Lang et al. 2019).

In intrusion detection systems available to date, there are two types of classification "detection-based" methods and data "source-based" methods (Heberlein et al. 1990). In the "detection-based" methods, the IDSs are splitters into misuse detection and abnormality detection. While in the date "source-based" methods the IDSs can be splitters into "host-based" and "network-based" methods. Figure (2) shows the classification method for the IDSs:

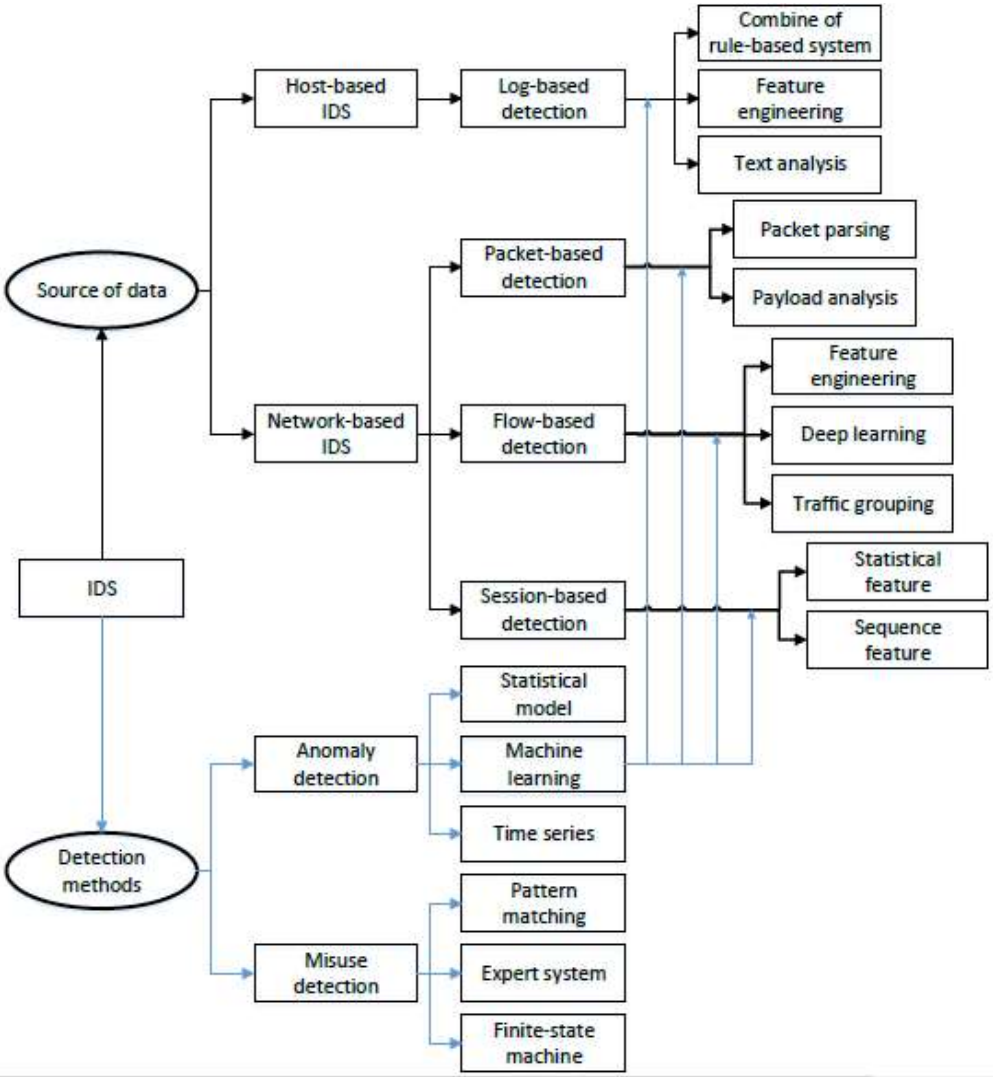

Figure 2: Classification Methods for IDs (Lang et al. 2019).

4. Related works

Karpagam et al. (2015) proposed a new approach for IDs based on pertinent features for the special attack. Here, IDs are done with the help of "supervised learning Neural Network (NN)", where the feature selection is performed with the help of information gain algorithm and GA. They test the relevant features using the "Multi-Layer Perceptron (MLP)" supervised NN. The experiment result shows that the performance for the IDS measures gets high accuracy when the selected features alone are used instead of all features.

Chowdhury et al. (2016) proposed a method to classify any thumping behavior in traffic of the network, using a hybrid approach of two ML algorithms. To evaluate the detection accuracy of their model, they used a "false positive rate", "false-negative rate" based on the time taken to detect the intrusion. The experiment results showed a higher detection rate and accuracy reach 98.76% and a lower "false-positive" rate reaches 0.09% and a "false-negative" rate reaches 1.15%. In contrast, the normal SVM-based scheme achieved detection accuracy of 88.03%, the "false-positive" rate reach 4.2% and the "false-negative" rate reaches 7.77%.

Almseidin et al. (2017) applied different experiments "J48, Random Forest, Random Tree, Decision Table, MLP, Naive Bayes, and Bayes Network" to assess various ML classifiers based on the "KDD intrusion dataset". To enhance the detection rate of the IDs system, they focused on the "false-negative" and "false-positive" performance metrics. The final results showed that the random forest classifier has the highest average accuracy rate while the decision table classifier achieved the lowest value of "false negative".

Alsawy et al. (2018) proposed a method that aims to classify the activities of networks as normal or abnormal, using different ML algorithms, "Random Forest (RF)", "Multi-Layer Perceptron (MLP)", and Library for "Support Vector Machine (LIBSVM)". They tested the proposed approach using a common dataset called "NSL-KDD" and applied the different ML algorithms over the dataset. They used the "Correlation Feature Selection (CFS)" as a Feature Selection algorithm, to drop out some irrelevant attributes from the dataset. The experiment results showed that the multilayer perceptron classifier accuracy was 95.7%, while the Random Forest's accuracy reaches 99.6%, and the LIBSVM classifier accuracy reaches 94.8%. Feature Selection “CFS” showed a low accuracy of 91.7%. However, with the LIBSVM, the accuracy rate increased to 97.2%.

Alomari et al. (2018) present a learning calculation for anomaly using a system interruption symmetry framework that utilizes choice tree computation. This method was designed to recognize assaults from average training and to sort out diverse kinds of interruptions. The machinery was based on modifying the weights of the dataset in light of eventualities, and thus keep splitting the dataset into sub-datasets until all sub-datasets have a place in a comparable class. The test that came about on the "KDD-99" bench-mark organized obstruction discovery data-set. Compared to other executing techniques, it turned out that the suggested algorithm resulted in a 98.5% detection rate.

Fattahi et al. (2018) present a hybrid method of the ANN and SVM. They compared the accuracy of their method with the varying results of other classifiers. The experiment results proceeded with a different selected network connected with the "NSL-KDD DARPA" dataset. The initial results showed the hybrid of "ANN" and "SVM" techniques for attack detection could be viable alternatives for future work.

Biswas (2018) proposed an intrusion detection method using ML with other feature selection techniques to select the important features from the original data of features and to sort them out by studying and analyzing the popular classifiers and methods of selection. They applied a five tucks cross-validation to get the results and figure out the accuracy of the "NSL-KDD" dataset. The experiment results showed that (1) the "K-NN" classifier has higher performance and efficiency than the other methods, and (2) the information gain ratio-based feature selection method is better off.

Hamdi et al. (2020) proposed an approach that contains two models for IDs and classification scheme "Trust-based" IDs and Classification System "TIDCS" and Trust-based IDs and Classification System- Accelerated "TIDCS-A" for a secure network. Where TIDCS aims to reduce the number of inserted using the proposed algorithm for feature selection. At first, the features are classifying randomly to raise the chance of making them participating in the generation of different classes, and classified based on their accuracy scores. Then, the high-ranked features will select to gain a classification for any received packet from the nodes in the network, which is saved as part of the node's past performance. "TIDCS" suggests a cyclic system cleansing where trust relationships between participant nodes are estimated and renewed periodically. "TIDCS-A" aims to make a dynamic algorithm to compute the exact time for nodes cleansing states and shackle the insinuation window of the nodes. The final classification decision for both methods is rated by combining the node's past behavior with the ML algorithm. The experiment results show the accuracy detection for the UNSW dataset equal 91% for TICDS, 83.47% by using online AODE, 88% for CADF, 90% for EDM, 90% for TANN, and 69.6% for NB.

Table 1: Research Summary

Author	Method	Accuracy	Weakness
Karpagamet al. (2015)	- "NN, GA"	- progress the Detection Rate (DTR) for the different types of attacks.	
Chowdhury et al. (2016)	- "SVM"	- Accuracy = 98.76%	- The complexity of selecting the appropriate features
Almseidin et al. (2018)	- "J48, Random Forest, Random Tree, Decision Table, MLP, Naive Bayes, and Bayes Network"	- Accuracy = 93.77%,	- no single ML algorithm can hold all the types of attacks.
Alsawy et al. (2018)	- "Random Forest (RF), Multi-Layer Perceptron (MLP), Library for Support Vector Machine (LIBSVM), Correlation Feature Selection"	- Accuracy = (99.6%).	- The weakness of the integration of multiple IDs at the runtime
Alomari et al. (2018)	- "Decision Tree (DT)"	- Accuracy = 98.5%.	- The lack of efficiency for detections using small datasets
Fattahi et al. (2018)	- "Artificial neural network (ANN), classifier and support vector machine (SVM)".	- increase the performance for supervised and unsupervised ML methods.	- the lack of classifying attacks based on various sources of data
(Biswas (2018	- "Naive Bayes, Support	- increase the performance for the IGR feature	- not success method for all features selection methods

	Vector Machine, Decision Tree, Neural Network, k- nearest neighbor algorithm (k-NN)"	selection and KNN.	(CFS)
Hamdi et al. (2020)	- TIDCS - TIDCS-A	- Accuracy = 90%	- A significant difference in results when applying the different methods

5 . Discussion

In this section, we compared the results for the different studies that work on intrusion detection, such as Alsawy et al. (2018), Kurniabudi et al. (2020), Vinayakumar et al. (2019), Panwar et al. (2019), Iman et al. (2018), and Yulianto et al. (2019). In Alsawy et al.'s (2018) study, they used three different classifiers (Multilayer perceptron classifier, Random Forest Classifier, and LIBSVM Classifier), while in the Kurniabudi et al. (2020) study, they used five classifiers (RF, BN, RT, NB, and J48). But in the Vinayakumar et al. (2019) study, they used 7 classifiers (LR, NB, KNN, DT, AB, RF, SVM-RBF). Iman et al. (2018) study used one classifier (AdaBoost), while Yulianto et al. (2019) used four classifiers (EFS, EFS with SMOTE, AdaBoost with PCA Feature, and AdaBoost with each of PCA Feature and SMOTE).

Finally, in the Panwar et al. (2019) study, they used (CfsSubset Attribute Evaluator, Classifier Subset Evaluator with Naive Bayes, CfsSubset Attribute Evaluator with J48, Classifier Subset, Evaluator with Decision Tree, CfsSubset Attribute Evaluator, Classifier Subset Evaluator with Naive Bayes, CfsSubset Attribute Evaluator with J48, and Classifier Subset Evaluator with Decision Tree). Table (2) shown the comparing results between the different studies with the proposed method:

Table 2: The Comparing results between the different studies

Study	Algorithm	Precision	Recall	F-Measure	Accuracy
(Alsawy et.Al 2018)	MLP	95.70%	97.10%	96.30%	93.80%
(Alsawy et.Al 2018)	RF	99.60%	99.60%	99.60%	99.60%
(Alsawy et.Al 2018)	LIBSVM	94.80%	95%	94.50%	97.20%
(Kurniabudi et.Al 2020)	RF	96.48%	99.90%	94.03%	99.40%
(Kurniabudi et.Al 2020)	BN	95.92	97.12%	94.03%	94.80%
(Kurniabudi et.Al 2020)	RT	98.97%	99.80%	94.03%	99.70%
(Kurniabudi et.Al 2020)	NB	70.84%	90%	94.50%	40.90%
(Kurniabudi et.Al 2020)	J48	98.88%	99.75%	95.04%	99.80%
(Vinayakumar et.Al 2019)	LR	68.50%	85.00%	75.80%	83.00%
(Vinayakumar et.Al 2019)	NB	30.00%	97.90%	45.90%	31.00%
(Vinayakumar et.Al 2019)	KNN	78.10%	96.80%	86.50%	91.00%
(Vinayakumar et.Al 2019)	DT	83.90%	96.50%	89.80%	94.00%
(Vinayakumar et.Al 2019)	AB	88.70%	92.80%	90.20%	94.00%
(Vinayakumar et.Al 2019)	RF	84.90%	96.90%	90.50%	94.00%
(Vinayakumar et.Al 2019)	SVM-rbf	99.30%	32.80%	49.30%	97.00%
(Panwar et.Al 2019)	CfsSubset Attribute Evaluator	99.24%	94.03%	94.03%	96.04%
(Panwar et.Al 2019)	Classifier Subset Evaluator With Naive Bayes	99.24%	99.68%	94.03%	98.96%
(Panwar et.Al 2019)	CfsSubset Attribute Evaluator with J48	99.24%	99.68%	94.03%	98.96%
(Panwar et.Al 2019)	Classifier Subset Evaluator With Decision Tree	99.24%	99.68%	94.03%	96.04%
(Panwar et.Al 2019)	CfsSubset Attribute Evaluator	99.57%	99.99%	95.04%	98.91%
(Panwar et.Al 2019)	classifier Subset Evaluator With Naive Bayes	99.59%	99.28%	97.15%	99.96%

(Panwar et.Al 2019)	CfsSubset Attribute Evaluator with J48	99.61%	99.99%	98.25%	99.98%
(Panwar et.Al 2019)	Classifier Subset Evaluator With Decision Tree	99.24%	99.68%	94.03%	96.04%
(Iman et.Al 2018)	AdaBoost	77%	88%	77%	77%
(Yulianto et.Al 2019)	EFS	85,15	94,92	89,77	81.47%
(Yulianto et.Al 2019)	EFS + SMOTE	81,83	100	90,01	81.83%
(Yulianto et.Al 2019)	AdaBoost + PCA Feature	81,49	99,93	89,78	81.47%
(Yulianto et.Al 2019)	AdaBoost + PCA Feature + SMOTE	81,69	95,76	88,17	81.47%

To compare these results for the different studies, we take the average of all studies for each of the precision, recall, and f-measure as shown in table (3), where the results showed a preference for the Panwar et al. (2019) study comparing to other studies:

Table 3: The average of results

Study	Precision	Recall	F-Measure	Accuracy
(Alsawy et.Al 2018)	96.70%	97.23%	96.80%	96.87%
(Kurniabudi et.Al 2020)	92.22%	97.31%	94.33%	86.92%
(Vinayakumar et.Al 2019)	76.20%	85.53%	75.43%	81.17%
(Panwar et.Al 2019)	99.37%	99.00%	95.07%	98.11%
(Iman et.Al 2018)	77%	88%	77%	77%
(Yulianto et.Al 2019)	82.54%	97.65%	89.43%	81.56%

5. Conclusion

To date, many works and researches have suggested several approaches for intrusion detection using different methods such as machine learning and deep learning. In this paper, we reviewed the most significant approaches that have been put forward in this field and compared the results for the latest studies for intrusion detection over the networks.

References

Raheem, A., and Alomari, E. (2018). An Adaptive Intrusion Detection System by using a decision tree. *Journal of AL-Qadisiyah for computer science and mathematics*. Vol.10 No.2.

Kumar, E. A., Kaur, J., and Kaur, I. (2016). Intrusion Detection System by Machine Learning Review. International Journal of Advanced Research, Ideas, and Innovations in Technology. 10.1007/978-81-322-2529-4_51, page-1.

Mohamed, H., Hefny, H., and Alsawy, A. (2018). Intrusion Detection System Using Machine Learning Approaches. Egyptian Computer Science Journal. Vol. 42 No.3.

Kaur, H., Singh, G., and Minhas, J. (2013). A Review of Machine Learning-based Anomaly Detection Techniques. International Journal of Computer Applications Technology and Research. Volume 2– Issue 2, 185 - 187, 2013.

Biswas, S.K. (2018). Intrusion Detection Using Machine Learning: A Comparison Study. International Journal of Pure and Applied Mathematics. Volume 118 No. 19.

Karpagam, S., Revathi, T., and Jayakumar, K. (2015). Intrusion Detection using Artificial Neural Networks with Best Set of Features. The International Arab Journal of Information Technology. Vol. 12, No. 6A.

Anderson, J.P. Computer Security Threat Monitoring and Surveillance, Technical Report; James P. Anderson Company: Philadelphia, PA, USA, 1980.

Lang, B., & Liu, H. (2019). Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. International Journal of Engineering & Technology. (3.24) (2018) 479-482.

Kumar, E. A., Kaur, J., and Kaur, I. (2016). Intrusion Detection System by Machine Learning Review. International Journal of Advanced Research, Ideas, and

Innovations in Technology. 10.1007/978-81-322-2529-4_51, page-1.

Juma, S., Muda, Z., Mohammad, M. A., and Yasin, W. (2015). Machine Learning Techniques for Intrusion Detection: A Review. Journal of Theoretical and Applied Information Technology. Vol.72 No.3.

Haq, N. F. et al. (2015). Application of Machine Learning Approaches in Intrusion Detection System: A Survey. (IJARAI) International Journal of Advanced Research in Artificial Intelligence. Vol. 4, No.3.

Akhilesh Kumar Shrivas, A. K. (2014). An Ensemble Model for Classification of Attacks with Feature Selection based on KDD99 and NSL-KDD Data Set. International Journal of Computer Applications, Volume 99 –No.15.

Heberlein, L.T., Dias, G.V., Levitt, K.N., Mukherjee, B., Wood, J., Wolber, D. (1990). A network security monitor. In Proceedings of the 1990 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA, USA, 7–9 May 1990; pp. 296–304.

Chowdhury, M. N., Ferens, H., and Ferens, M. (2016). Network Intrusion Detection Using Machine Learning. Int'l Conf. Security and Management.

Almseidin, M., Alzubi, M., Kovacs, S., and Alkasassbeh, M. (2018). Evaluation of Machine Learning Algorithms for Intrusion Detection System. International Symposium on Intelligent Systems and Informatics (SISY).1949-0488.

Fattahi, J., Omrani, T., Dallali, A., &Rhaimi, B. (2018). The fusion of ANN and SVM Classifiers for Network Attack Detection. arXiv:1801.02746v2 [cs.CR] 10 Jan 2018.

HAMDI, A.M., GUIZANI, M., MOHAMED, A., HAMILA, R., ERBAD, A., and CHKIRBENE, Z. (2020). TIDCS: A Dynamic Intrusion Detection and Classification System Based Feature Selection. in IEEE Access, vol. 8, pp. 95864-95877, 2020, DOI: 10.1109/ACCESS.2020.299493.