INTRUSION DETECTION BASED ON DEEP LEARNING TECHNIQUES IN COMPUTER NETWORKS

L K Suresh Kumar
University College of Engineering, Osmania University, India

ABSTRACT

Security Breaches in computer networks have increased a lot in the last decade due to the profitable underground cybercrime economy. Many researches have been working on finding efficient techniques for detecting intrusions. Many surveys were present on different Machine Learning and Deep Learning Techniques in the last decade. Solutions proposed for dealing with network intrusions can be broadly classified as signature based and anomaly based. In this paper, a critical survey of Machine Learning (ML) and Deep learning (DL) techniques presented in the literature in the last ten years is presented. This survey would serve as a supplement to other general surveys on intrusion detection as well as a reference to recent work done in the area for researches working in ML and DL based intrusion detection systems. Some open issues are also discussed that are needed to be addressed.

KEYWORDS: Computer security, Deep Learning, Intrusion detection, Machine Learning, Security Breaches

1. INTRODUCTION

With recent advancements in technology and availability of sophisticated tools, intrusions in computer networks have increased significantly. Intrusion is an attempt to compromise CIA (Confidentiality, Integrity, and Availability) or to bypass the security mechanisms of a computer or a network. Intrusions or security breaches incurs a huge loss to the companies. Table 1. describes the impact of intrusions on various companies. Intruders are using sophisticated (highly developed) tools to breach security. It's very hard to detect those security breaches with the old traditional tools for intrusion detection. Hence, there is an alarming need to develop sophisticated intrusion detection as well as Intrusion Prevention Systems. The software or hardware which continuously monitor the network traffic is efficient in detecting security breaches. Many researchers have started working on Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS). One such research started in 1972[1] when James Anderson published his report on the need for detecting breaches in computer systems[2].

Table 1. Impact of Security Breaches on various companies

Company	Date	Impact
Yahoo	August 2013	3 billion accounts
LinkedIn	June 2021	700 million users

Sina Weibo	March 2020	538 million accounts
Facebook	April 2019	533 million users
Yahoo	2014	500 million accounts
NetEase	October 2015	235 million user accounts
LinkedIn	June 2012	165 million users
Adobe	October 2013	153 million user records

Intrusion Detection is the process of monitoring the events occurring in a computer system or network and analysing them for signs of intrusion. IDS and IPS systems are classified into many types based on how these systems collect, process and act upon the events that are related to security breaches. There are mainly three types namely centralized, distributed or hybrid. These systems differ in terms of cost, performance and other measures. Intrusion Detection Systems are mainly classified as Signature-based and Anamoly-based. Signature based IDSes use the signature of known attacks to determine intrusions. Whereas, Anomaly based IDSs collect the data relating to behaviour of legitimate users and then current observed behaviour is analyzed to determine if those are that of authorized users or malicious users.

There are different models used for Intrusion Detection. The Fields that are related to these models are Artificial Intelligence (AI), Machine Learning (ML), Neural Networks (NN) or Artificial Neural Networks (ANN), Deep Learning (DL). These fields are interrelated to each other as shown in the Fig1. Artificial Intelligence is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. Machine Learning is a subfield of Artificial Intelligence and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. NNs are the network of neurons which pass data through several layers of the interconnected neurons. Haykin.S [3] defines NN as “Massively parallel combination of simple processing units which can acquire knowledge from the environment through a learning process and store the knowledge in its connections”.

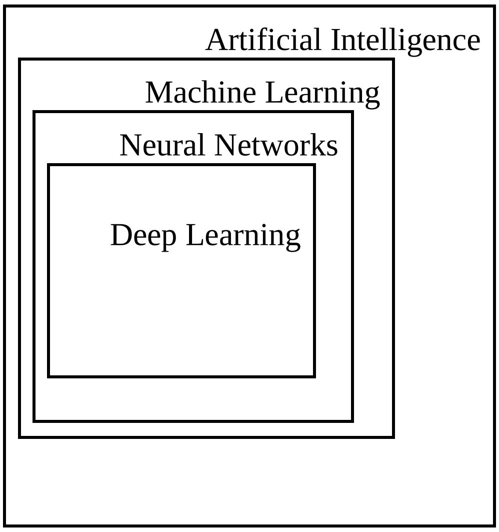

Fig 1 Interrelation between different fields

There are different ML based approaches and DL based approaches used for Intrusion Detection. Some of the ML based approaches are Artificial Neural Networks, Association Rules, Fuzzy Association Rules, Bayesian Network, Clustering, Decision Trees, Ensemble Learning, Evolutionary Computation, Hidden Markov Models, Inductive Learning, Naïve Bayes, K-Nearest Neighbours[5] and Support Vector Machine[4]. DL (Deep Learning) based approaches used for Intrusion detection are Deep neural network, Feed forward deep neural network[6], Recurrent neural network, Convolutional neural network, Deep auto-encoder[7]. In this paper, the main focus is to present a critical survey on Intrusion detection approaches based on ML as well as DL presented in literature. The paper is organized as follows: In section2, Required background Information is presented. In section3, a survey on ML and DL based intrusion detection techniques is presented. In section 4 , related work is discussed. In section 5, observations and open issues about the survey are addressed. Section 6 concludes the paper.

2. BACKGROUND

ML Algorithms are broadly classified as (i)Supervised (ii)Unsupervised (iii)Reinforcement learning. These are the three pillars of Machine Learning. Supervised learning algorithms use labelled datasets for training the model, which can then be used for purposes such as: Classification and Regression. In unsupervised learning, the training data are not labelled; the learning algorithm used tries to group / classify the training data based on some grouping techniques. Reinforcement learning is the third main class of machine learning algorithms which aims to find the middle ground between exploration of the data, such as unsupervised learning, and the usage of that knowledge, such as supervised learning. Unlike supervised learning it does not require a labelled dataset, and unlike unsupervised learning it does not focus solely on finding patterns without an intended application. Instead, it seeks to maximise the reward function in the long term even if that means experiencing negative reward, otherwise known as regret, in the short term for a overall experience. Signature based IDEes generally use supervised learning algorithms whereas anomaly-based IDEes use unsupervised learning algorithms.

Binary classification is the basic classification type where the number of classification labels is two. For intrusion detection, binary classification uses the labels normal and attack (or anomalous, abnormal, etc). Under Multiclass classification, the number of labels used for classifying data can be more than two. For example, for intrusion detection, data can be labelled as normal, Denial of Service Attack (DoS), User to Root Attack (U2R), Remote to Local Attack (R2L), Probing Attack, etc. Next, some types of algorithms used for classifying data are discussed.

2.1. Machine Learning Techniques

2.1.1. K-NN algorithm based classification

K-Nearest Neighbours is one of the simplest Supervised Machine Learning algorithm mostly used for classifying a data point based on how its neighbours are classified. KNN is based on feature similarity [8]. KNN stores all available cases and classifies new cases based on a similarity measure. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. The basic idea behind K-NN algorithm is developed by Fix et al[9] and later generalized by Cover et al [5]. For example, if training data are classified into two types A and B. Data of type A are square in shape and the data of type B are round in shape. When a new data point which is square in shape needs to be classified, K-NN selects K nearest points and calculates the Euclidean distances between the new data point and the chosen K neighbours. Then, using some algorithm, it decides whether to classify the new data point as type A or type B.

2.1.2. SVM Algorithm based Classification

Support Vector Machine or SVM [10] is one of the most popular supervised learning algorithms, which is used for Classification as well as Regression problems. SVM chooses the extreme points / vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Suppose we see a strange dog that also has some features of foxes, so if we want a model that can accurately identify whether it is a dog or fox, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of dogs and foxes so that it can learn about different features of dogs and foxes, and then we test it with this strange creature. So, as support vector creates a decision boundary between these two data (dog and fox) and choose extreme cases (support vectors), it will see the extreme case of dog and fox. On the basis of the support vectors, it will classify it as a dog. There are mainly two types of SVM namely Linear SVM and Non-Linear SVM. Linear SVM is used for linearly separable data and Non-Linear SVM is used for non-linearly separated data.

2.2. Deep Learning Techniques

There are different deep learning-based approaches which are used for intrusion detection.

2.2.1. Autoencoders

It is an unsupervised learning technique. Autoencoders map the input data to output data by encoding the input data. It is a data compression algorithm. Autoencoders consists of mainly two parts namely encoder and decoder. Encoder takes the input and encodes it. For an input y, the encoded function is f(y). There lies an internal hidden layer between encoder and decoder which learns the coding from encoder. It has a function h, where h=f(y). Decoder function reconstructs the input from the hidden layer. If the decoder function is g, then reconstruction r=g(f(y)).

2.2.2. Recurrent neural networks

RNN [11] works on the principle of saving the output of a particular layer and feeding it back to the input to predict the output of the layer. RNN are an alternative to feed-forward neural network [12] because the latter cannot handle sequential data, it considers only the current input and it cannot memorize previous inputs.

2.3. Datasets used for training ML and DL Models

2.3.1. KDD99

Knowledge Discovery in Databases[13] was published in 1999. Stolfo et al[14],[15] prepared this dataset. It was prepared based on DARPA98[16]. The dataset consists of records and each record has 41 features. Records are tagged either as normal or an attack. This Dataset is imbalanced as the training and testing datasets of this dataset have 78% and 75% duplicate records. Table 2. describes the data distribution for each class of KDD99[17].

Table 2. Class Distribution of the KDD99 dataset

Class	Training	(%)	Testing	(%)
Normal	972781	19.85	60593	19.48
DoS	3883390	79.27	231455	74.41
Probe	41102	0.83	4166	1.33
U2R	52	0.001	245	0.07
R2L	1106	0.02	14570	4.68

2.3.2. NSL-KDD

The other dataset used for Intrusion detection is Network Socket Layer-Knowledge Discovery in Databases(NSL-KDD)[18]. Travellaee et al.[15] prepared this dataset on the basis of KDD99 dataset. This dataset does not contain any duplicate records. However, the class distribution is imbalanced like KDD99 dataset. Table 3 describes the class distribution of this dataset.

Table 3.Class distribution of the NSL-KDD dataset

Class	Training20	(%)	Training+	(%)	Testing	(%)
Normal	13449	53.3	67343	53.5	9711	43.1
DoS	9234	36.7	45927	36.4	7458	33.1
Probe	2289	9.1	11656	9.3	2421	10.7
U2R	11	0.04	52	0.041	67	0.3
R2L	209	0.83	995	0.78	2887	12.8

2.3.3. UNSW-NB15

Moustafa et al. [19] proposed this dataset. This dataset capture the network traffic with the help of tcp dump tool. This dataset consists of 49 features and 2,540,044 records. This dataset is also not balanced. The dataset contains data related to nine attack classes namely Fuzzers, Analysis Backdoor, DOS, Exploit, Generic, Reconnaissance, Shellcode and Worm. Table 4. describes the data distribution for various attack classes in the UNSW-NB15 dataset.

Table 4. Data distribution for various attack classes in the UNSW-NB15 dataset

Class	Training	(%)	Testing	(%)
Normal	11200	32	7400	45
Generic	8000	23	3774	23
Exploits	6679	19.05	2226	13.5
Fuzzers	3637	10.37	1212	7.36
DoS	2453	6.99	818	4.97
Reconnaissance	2098	5.98	699	4.25
Analysis	400	1.14	135	0.82
Backdoor	349	1	117	0.71
Shellcode	227	0.65	76	0.46
Worms	26	0.07	9	0.05

2.4. Performance metrics used for evaluating intrusion detection systems

True Positive Rate (TPR): Probability that an actual intrusion will be declared as an intrusion by the Intrusion Detection System.

TPR=TP/TP+FN ….(1)
Where, TP - Number of positive events detected by an IDS
FN - Number of false negative events detected by an IDS

True Negative Rate (TNR): Probability that a non-intrusive action will be declared as a non-intrusive action by the Intrusion detection System.

TNR=TN/TN+FP ….(2)
Where, TN - Number of true negative events detected by an IDS
FP - Number of false positive events detected by an IDS

False Positive Rate (FPR): Probability that an IDS will detect an action as intrusion while it is actually not an intrusion.

FPR=FP/FP+TN ….(3)
Where, TN-Number of true negative events detected by an IDS
FP-Number of false positive events detected by an IDS

False Negative Rate (FNR): Probability that a non-intrusive action/event will be detected as an intrusive action/event by the IDS

FNR=FN/FN+TP ….(4)
Where, TP - Number of true positive events detected by an IDS
FN - Number of false negative events detected by an IDS

Accuracy: It is the proportion of events that are correctly predicted by the algorithm.

Accuracy=TP+TN/TP+TN+FP+FN ….(5)

Precision: It is the proportion of events that are predicted by the algorithm as intrusion are actually intrusions.

Precision=TP/TP+FP ….(6)

Recall: It is the proportion of actual intrusions that were predicted as intrusions by the algorithm.

Recall=TP/TP+FN ….(7)

F1-Score: It is the harmonic mean of Precision and Recall

F1-Score=2*(Precision*Recall/Precision+Recall) ….(8)

2.4.1. Confusion Matrix

It is a table layout of experimental results of an algorithm.It helps in understanding the performance of an algorithm. Confusion matrix looks like Table 5.

Table 5

TP	FP
FN	TN

3. Intrusion detection techniques based on Machine Learning and Deep Learning

In this section, critical survey of the papers on Intrusion detection that use ML and DL techniques is presented. These papers used both binary as well multi class classification.

3.1. Intrusion detection techniques based on Machine Learning

Elbasiony et al. [20] proposed a hybrid Model which consists of two parts: misuse (signature-based) detection part and anomaly detection part. It has two phases namely online and offline. Online Phase is responsible for the misuse detection part. If no match is found, It sends the network traffic to Anomaly detection part. Weighted K-means algorithm is used in anomaly detection part. This hybrid model achieved 98.3% accuracy with 1.6% false positive rate on the KDD99 dataset.

Ever et al. [21] has proposed three Machine Learning models namely Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree (DT). The goal of the study was to find the best technique among the three. Two experiments were performed by using 60% and 70% of the dataset NSL-KDD for training and the rest of the dataset for testing. The experiments show that DT achieved 98.84% training accuracy. However, the accuracy of the other two models are not mentioned.

Horng et al.[22] used Support Vector Machine for Intrusion detection. Hierarchical clustering algorithms are used which reduce the size of training data because SVM cannot accept large input in it’s training phase. The outputs of the experiments show that this technique achieved 95.72% accuracy with 0.7% false positive rate.

Moualla et al.[23] used ML techniques to develop an Intrusion detection system. Synthetic minority oversampling (SMOTE) algorithm is used for balancing the dataset which is followed by the pre-processing stage where simple data cleaning and one-hot-encoding is performed. The normalization of data is done based on z-score. Extremely randomized trees classifier is used for feature selection and ELM classifier for each class type. This approach achieved an overall accuracy of 98.43%.

3.2. Intrusion detection techniques based on Deep Learning

Zhang et al. [24] proposed a technique based on Genetic Algorithm (GA) and Deep Brief Network (DBN) for addressing security in IoT. Genetic algorithm was used for optimizing the number of hidden layers and the number of neurons in a layer for each type of attack. They used NSL-KDD dataset to evaluate their technique. Before applying the technique normalization of the dataset is done using min-max normalization. This method has higher accuracy on classes like U2R compared to other compared methods.

Shone et al. [25] proposed a technique where he used two layers of Non-symmetric Deep Auto Encoder (NDAE) for unsupervised feature learning. NDAE does not contain a decoder. Random Forest (RF) was used to do the final classification of the network data into normal and attack. The authors used datasets like NSL-KDD and KDD99 for evaluating their model 10-fold cross validation was performed to overcome over-fitting and under-fitting problem. False alarm rate was very highin some attack classes as the datasets used contained imbalanced data.

Salama et al. [26] introduced a three step approach for Intrusion detection in computer networks. The first step includes pre-processing of data by changing symbolic features into numeric values. In the second step, they used DBN with two layers of Restricted Boltzmann Machine (RBM) to reduce the dimension of features. Back propagation algorithm was used for DBN. The first layer consists of 41 inputs. It was reduced to 13 and was fed as input to the second layer. The output of the second layer consists of 5 features. In the third step, SVM was used for classification purposes. The accuracy of their hybrid model was better than DBN and SVM.

Javaid et al. [27] proposed a model based on Autoencoder for feature representation and feature learning. Soft-max regression was used for the classification. In the pre-processing step, Transformation of categorical features into continuous features is done. They also normalized the data using min-max method. They performed two types of evaluations. In order to do cross validation, they used training data for both training and testing. In the second approach, they used different datasets for testing and training.

Yin et al. [11] introduced a two step deep learning approach for Intrusion detection. one hot encoding was used in the pre-processing step for transforming the categorical data into numerical values. After that, the dataset was normalized using min max method. For classification, RNN with forward propagation and backward propagation was used. Cross-entropy was used to calculate the difference between output values, produced by the forward propagation and true value. Both binary and multi-class classification were performed in this approach. Performance of this approach using KDD Test 21 was around 64%.

Ge et al. [28] introduced a deep learning based approach for detecting intrusions in IoT. They used FNN to detect intrusions. Information from header fields in IP packets is used to extract features. After extracting these features, they fed all the training data to the FNN to do the classification. Both binary and multi-class classification are performed. They used BoT IoT dataset, which was created in the cyber range lab of the UNSW Institute for Cyber Security. This method achieved higher accuracy than Support Vector Classifier (SVC).

Vinayakumar et al. [29] introduced a hybrid DNN framework for detecting intrusions, called scale-hybrid-IDS-AlertNet. This model can detect both host and network level intrusions. Different datasets like KDDCup99, NSL-KDD, Kyoto, UNSW-NB15, WSN-DS and CICIDS 2017 were used to test their model. Overall performance of their model is best among all the competitor models compared.

4. Related work

In this section, the survey papers on Intrusion detection systems, published over the last decade are briefly discussed. Ganapathy et al. [30] present a survey on feature selection and classification techniques. They also present two new algorithms - an attribute selection based feature selection algorithm and a rule-based multiclass support vector machine algorithm. Mitchell et al. [31] present a survey of 28 intrusion detection techniques for cyber-physical systems. They discuss the performance metrics used in IDSes, the characteristics of the IDSes and their merits and demerits.

Butun et al. [32] present a survey of IDSes for wireless sensor networks. Milenkoski et al. [33] present a survey of the common practices used for evaluating IDSes. They categorized the IDS based on monitored platform, attack detection method and deployment architecture. Then, they survey the evaluation approaches and methods used by these IDSes with respect to workload, metrics used and measurement methodology used. Vasilomanolakis et al. [34] present a detailed framework of requirements and building blocks for Collaborative Intrusion Detection Systems (CIDSes). They also present a critical analysis of CIDSes presented in the literature with respect to their framework.

Buczak et al. [35] present an article on the history of the evolution of ML and DM techniques. They also mentioned the differences, similarities, and various datasets used by these techniques. They also present about ML / DM methods in cybersecurity like Artificial Neural Networks, Association Rules, Fuzzy Association Rules, Bayesian Network, Clustering, Decision Trees, etc. They also discussed the computational complexity of ML and DM methods and related open issues.

Liu et al. [36] present a survey on different security threats against algorithms such as Naive Bayes, DT, SVM etc. Benkhelifa et al. [37] present a survey on Intrusion detection techniques presented in the literature for IoT. They also present an architecture for IDSes which are suitable for IoT. Nisioti et al . [38] present a survey on unsupervised and hybrid Machine Learning Approaches for Intrusion detection.They analyzed the algorithms and identified their drawbacks. They also identified some new classes of attacks that do not belong to any of the known classes and suggested methods to detect them.

Resende et al. [39] present a survey of IDSes using Random Forest based methods for classification, feature selection etc. Khraisat et al. [40] present a survey on some recent works of IDSes and also discuss some datasets used for evaluating them. Liang et al. [41] present a survey on advantages and disadvantages of using Machine Learning in designing Intrusion detection systems. Kiennert et al. [42] present a survey of Intrusion detection techniques that use game theory for analysing data and detecting intrusions. Al-Garadi et al.[43] present a survey on IDSes for Internet of Things (IoT). Wu et al. [44] present a survey on IDSes proposed in the literature for in-vehicle network (IVN).

5. Discussion and open issues

In this section, some open issues that are still to be resolved are discussed.

Parallelizing IDSes and distributed IDSes

Centralized Intrusion detection systems are not scalable and are suitable to only small networks. But, ML based centralized IDSes which use parallel processing are efficient in early detection of network based attacks. Distributed and collaborative IDSes are scalable but they may not detect all intrusions unless all events are collected and processed at a central location.

Dealing with data imbalance problem

All methods discussed so far cannot handle imbalanced datasets. One method which is robust in dealing with imbalanced datasets is Random Forest. More research work has to be done in finding efficient methods for dealing with data imbalance problem.

Labelling of training data

Supervised algorithms of Machine Learning need Labelled data for training. Effective methods for determining labels for training data are needed. More research work has to be done in finding efficient methods for determining labels for training data.

Security of machine learning algorithms.

Machine Learning Algorithms are susceptible to attacks. Attackers could inject malicious data into training datasets. These types of attacks are possible especially against ML algorithms that dynamically update their training datasets.

Performance metrics

Detection latency is an important metric for evaluating IDSes. Timely detection of intrusions will help in reducing recovery time and recovery overhead. So, More research work has to be done in finding techniques that have good Detection latency.

6. Conclusion

In this paper, a critical survey of research work on intrusion detection using ML techniques and DL techniques over the last decade are discussed. Some of the open issues that still remain to be addressed are also discussed. This survey is complementary to other existing surveys on intrusion detection and will serve as a supplement to other surveys. It will also serve as ready reference to researchers working on intrusion detection using ML and DL techniques.

References

[1] Anderson J.P.**Computer Security Technology Planning Study, Vol. 2** James P. Anderson & Co., Fort Washington, PA (1972)
[2] Bridges R., Glass-Vanderlan T., Iannacone M., Vincent M., Chen Q.**A survey of intrusion detection systems leveraging host data** ACM Comput. Surv., 52 (6) (2020), pp. 1-35
[3] Haykin S. **Neural Networks: A Comprehensive Foundation** Prentice Hall (1999)
[4] Noble W.S.**What is a support vector machine?** Nature Biotechnol., 24 (12) (2006), pp. 1565-1567

[5] Cover T.M., Hart P.E.**Nearest neighbor pattern classification** IEEE Trans. Inform. Theory, 13 (1) (1967), pp. 21-27, 10.1109/TIT.1967.1053964
[6] Hornik K., Stinchcombe M., White H.**Multilayer feed forward networks are universal approximators** Neural Netw., 2 (5) (1988), pp. 359-366, 10.1016/0893-6080(89)90020-8
[7] Abusitta A., Bellaiche M., Dagenais M., Halabi T. **A deep learning approach for proactive multi-cloud cooperative intrusion detection system.** Future Gener. Comput. Syst., 98 (2019), pp. 308-318, 10.1016/j.future.2019.03.043
[8] Jain A.K., Dubes R.C. **Algorithms for Clustering Data** Prentice Hall, Englewood Cliffs, NJ, USA (1988)
[9] Fix E., Hodges J. **Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties: Tech. Rep.** USAF School of Aviation Medicine, Randolph Field, Texas (1951)
[10] Joachims T. **SVM light: Support vector machine** (2008)
[11] Yin C., Zhu Y., Fei J., He X. **A deep learning approach for intrusion detection using recurrent neural networks** IEEE Access, 5 (2017), pp. 21954-21961
[12] Huang G.-B., Zhu Q.-Y., Siew C.-K.**Extreme learning machine: a new learning scheme of feed forward neural networks** Proceedings of 2004 IEEE International Joint Conference on Neural Networks, vol. 2, IEEE (2004), pp. 985-990
[13] **KDD cup 1999 data** (2021) http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, October 2007. (Accessed 21 August 2021)
[14] Stolfo S.J., Fan W., Lee W., Prodromidis A., Chan P.K.**Cost-based modeling for fraud and intrusion detection: Results from the JAM project** Proceedings DARPA Information Survivability Conference and Exposition, vol. 2,DISCEX'00, IEEE (2000), pp.130-144
[15] Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. **A detailed analysis of the KDD CUP 99 data set** Proceedings of 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, IEEE (2009), pp. 1-6
[16] Lippmann R.P., Fried D.J., Graf I., Haines J.W., Kendall K.R., McClung D., Weber D., Webster S.E., Wyschogrod D., Cunningham R.K., *et al.* **Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation** Proceedings DARPA Information Survivability Conference and Exposition, vol.2, DISCEX'00, IEEE (2000), pp12- 26
[17] Özgür A., Erdem H.**A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015** Peer J Preprints, 4 (2016), p. e1954v1
[18] **NSL-KDD dataset** (2021) https://www.unb.ca/cic/datasets/nsl.html. (Accessed 21 August 2021)
[19] Moustafa N., Slay J. **UNSW-NB15: A comprehensive data set for network intrusion detections systems (UNSW-NB15 network data set)** Proceedings of 2015 Military Communications and Information Systems Conference, MilCIS, IEEE (2015), pp. 1-6
[20] Elbasiony R.M., Sallam E.A., Eltobely T.E., Fahmy M.M. **A hybrid network intrusion detection framework based on random forests and weighted k-means** Ain Shams Eng. J., 4 (4) (2013), pp. 753-762
[21] Ever Y.K., Sekeroglu B., Dimililer K. **Classification analysis of intrusion detection on NSL-KDD using machine learning algorithms.** Proceedings of International Conference on Mobile Web and Intelligent Information Systems, Lecture Notes in Computer Science, vol. 11673, Springer, Cham (2019)
[22] Horng S.-J., Su M.-Y., Chen Y.-H., Kao T.-W., Chen R.-J., Lai J.-L., Perkasa C.D. **A novel intrusion detection system based on hierarchical clustering and support vector machines.** Expert Syst. Appl., 38 (1) (2011), pp. 306-313, 10.1016/j.eswa.2010.06.066

[23] Moualla S., Khorzom K., Jafar A. **Improving the performance of machine learning based network intrusion detection systems on the UNSW-NB15 dataset** Comput. Intell. Neurosci., 2021 (2021)

[24] Zhang Y., Li P., Wang X. **Intrusion detection for IoT based on improved genetic algorithm and deep belief network** IEEE Access, 7 (2019), pp. 31711-31722, 10.1109/ACCESS.2019.2903723

[25] Shone N., Ngoc T.N., Phai V.D., Shi Q. **A deep learning approach to network intrusion detection.** IEEE Trans. Emerg. Top. Comput. Intell., 2 (1) (2018), pp. 41-50, 10.1109/TETCI.2017.2772792

[26] Salama M.A., Eid H.F., Ramadan R.A., Darwish A., Hassanien A.E. **Hybrid intelligent intrusion detection scheme** Soft Computing in Industrial Applications, Springer (2011), pp. 293-303

[27] Javaid A., Niyaz Q., Sun W., Alam M. **A deep learning approach for network intrusion detection system.** Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies (Formerly BIONETICS) (2016), pp. 21 26

[28] Ge M., Fu X., Syed N., Baig Z., Teo G., Robles-Kelly A. **Deep learning-based intrusion detection for IoT networks** Proceedings of IEEE 24th Pacific Rim International Symposium on Dependable Computing, PRDC, IEEE, Kyoto, Japan (2019), 10.1109 / PRDC47002.2019.00056

[29] Ge M., Fu X., Syed N., Baig Z., Teo G., Robles-Kelly A. **Deep learning-based intrusion detection for IoT networks** Proceedings of IEEE 24th Pacific Rim International Symposium on Dependable Computing, PRDC, IEEE, Kyoto, Japan (2019), 10.1109 / PRDC47002.2019.00056

[30] Ganapathy S., Kulothungan K., Muthurajkumar S., Vijayalakshmi M., Yogesh P., Kannan A. **Intelligent feature selection and classification techniques for intrusion detection in networks: a survey** EURASIP J. Wireless Commun. Networking, 2013 (1) (2013), pp. 1-16, 10.1186/1687-1499-2013-271

[31] Mitchell R., Chen I.R. **A survey of intrusion detection techniques for cyber-physical systems** ACM Comput. Surv., 46 (4) (2014), 10.1145/2542049 Article No.: 55

[32] Butun I., Morgera S.D., Sankar R. **A survey of intrusion detection systems in wireless sensor networks.** IEEE Commun. Surv. Tutor., 16 (1) (2014), pp. 266-282

[33] Milenkoski A., Vieira M., Kounev S., Avritzer A., Payne B.D. **Evaluating computer intrusion detection systems: A survey of common practices** ACM Comput. Surv., 48 (1) (2015), pp. 1-41, 10.1145/2808691

[34] Vasilomanolakis E., Karuppayah S., Mühlhäuser M., Fischer M. **Taxonomy and survey of collaborative intrusion detection** ACM Comput. Surv., 47 (4) (2015), pp. 1-33

[35] Buczak A.L., Guven E. **A survey of data mining and machine learning methods for cyber security intrusion detection** IEEE Commun. Surv. Tutor., 18 (2) (2016), pp. 1153-1176, 10.1109/COMST.2015.2494502

[36] Liu Q., Li P., Zhao W., Cai W., Yu S., Leung V.C.M.**A survey on security threats and defensive techniques of machine learning: A data driven view** IEEE Access, 6 (2018), pp. 12103-12117, 10.1109/ACCESS.2018.2805680

[37] Benkhelifa E., Welsh T., Hamouda W. **A critical review of practices and challenges in intrusion detection systems for IoT: Toward universal and resilient systems** IEEE Commun. Surv. Tutor., 20 (4) (2018), pp. 3496-3509, 10.1109/COMST.2018.2844742

[38] Nisioti A., Mylonas A., Yoo P.D., Katos V. **From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods** IEEE Commun. Surv. Tutor., 20 (4) (2018), pp. 3369-3388
[39] Resende P., Drummond A. **A survey of random forest based methods for intrusion detection systems** ACM Comput. Surv., 51 (3) (2018), pp. 1-36
[40] Khraisat A., Gondal I., Vamplew P., Kamruzzaman J.**Survey of intrusion detection systems: techniques, datasets and challenges** Cybersecurity, 2 (2019), p. 20
[41] Liang F., Hatcher W.G., Liao W., Gao W., Yu W. **Machine learning for security and the internet of things: The good, the bad, and the ugly** IEEE Access, 7 (2019), pp. 158126-158147, 10.1109/ACCESS.2019.2948912
[42] Kiennert C., Ismail Z., Debar H., Leneutre H. **A survey on game-theoretic approaches for intrusion detection and response optimization** ACM Comput. Surv., 51 (5) (2019), pp. 1-31
[43] Al-Garadi M.A., Mohamed A., Al-Ali A.K., Du X., Ali I., Guizani M. **A survey of machine and deep learning methods for internet of things (IoT) security** IEEE Commun. Surv. Tutor., 22 (3) (2020), pp. 1646-1685,10.1109/COMST.2020.2988293
[44] Wu W., Li R., Xie G., An J., Bai Y., Zhou J., Li K. **A survey of intrusion detection for in-vehicle networks** IEEE Trans. Intell. Transp. Syst., 21 (3) (2020), pp. 919-933
[45] Viswanatha Reddy, Dr. Elango NM and Dr. C Kishor Kumar Reddy, “ Internet of Things Based Early Detection of Diabetes Using machine Learning Algorithms:DPA”, International Journal of Innovative Technology and Exploring Engineering, 2019

[46] Viswanatha Reddy, Dr. Elango NM and Dr. C Kishor Kumar Reddy, “Diabetes KaggleDataset Adequacy Scrutiny using Factor Exploration and Correlation”, International Journal of Recent Technology and Engineering, 2019

[47] Viswanatha Reddy, Dr. Elango NM and Dr. C Kishor Kumar Reddy, “Diabetes Diagnosis and Prognosis using Machine Learning Approaches: A Survey, “ International Journal of Advanced Science and Technology, 2019

[48] Kishor Kumar Reddy C and Vijaya Babu B, “ISPM-OC: Improved Snow Prediction Model Using Optimal k-Means Clustering and Decision Tree to Nowcast Snow/No-Snow”, International Journal of Control Theory and Applications , 2017

(http://serialsjournals.com/abstract/41550_14-c._kishor_kumar_reddy.pdf)

[49] Kishor Kumar Reddy C and Vijaya Babu B, “ISLIQ-OC: Improved Supervised Learning in Quest using Optimal k-means Clustering Mechanism to Nowcast Snow/No-Snow”, International Journal of Control Theory and Applications , 2017.