Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EVALUATION OF DEEP LEARNING AND MACHINE LEARNING ALGORITHMS IN INTRUSION DETECTION SYSTEMS

SAWSAN ALSHATTNAWI
Computer Science Department
Yarmouk University, Jordan
Email: sawsan_kh@yu.edu.jo

ABSTRACT

The advances in using technologies and the Internet increase the spread of data over the network. New attacks may be generated to threaten network security and data integrity, in addition, intruders always exist and we can't ignore their presence. Many tools and algorithms have been used to create many intrusion detection systems. A Network Intrusion Detection System (NIDS) is one of these tools designed to detect intrusion and stop it. Machine Learning (ML) and Deep Learning (DL) algorithms are used to build powerful NIDS. Deep learning proved its success by giving high accuracy and best performance compared with machine learning algorithms. In this paper, we will compare the accuracy of the used approaches by reviewing the literature and focusing on the research that used a famous and well-known dataset called KDD Cup'99. The paper will, in the end, summarize the issues and challenges in NIDS. The study summarizes the main issues and challenges of using the machine and deep learning and advice the researchers to start new approaches such as the deep forest over more recent datasets.

Keywords: *Network Intrusion Detection systems, Machine Learning, Deep Learning, KDD Cup'99.*

1. INTRODUCTION

Intrusion detection systems (IDS) are devices or applications designed to detect intrusions that occur on a computer network to maintain the security, integrity, and confidentiality of the network system. With the rapid proliferation of the Internet and the huge amount of data stored and transferred over the network, the IDS must be powerful and improve detection accuracy [1][2]. Artificial Intelligence (AI) approaches are used to improve the accuracy of NIDS. Machine Learning and Deep Learning approaches under the umbrella of AI are used to build a powerful NIDS.

Intrusion detection systems are divided into four types, the first two types depend on the location of deployment of IDS, host or network, the host scans the internal data of the local computer system, and the network examines the data exchanged between the computer systems, the second two types depend on the way of attack detection; signature or anomalies based approaches. The signature-based uses certain patterns such as the sequence of bytes in network traffic and then compares it to an existing database of signatures, and the anomalies-based which compares the behavior of a network with a trustworthy activity model, this type may detect previously unknown attacks but the problem, it can classify the previously unknown trustworthy activity as malicious activity[2][3].

IDS handles very large data that may contain inappropriate, untrustworthy, and often frequent features. These features help to quickly detect and identify an attack or malicious procedure. The presence of a feature selection system in IDS helps to select the most relevant or trustworthy features that may greatly aid the rating process. Feature selection is one of the machine learning techniques and it is implemented using many algorithms such as statistical analysis and swarm intelligence.

The researchers have built several systems for cyber security in different network and computing environments to detect intrusions, such as in the wireless sensor network [6], e-healthcare system [7], cloud computing [8], fog computing [9], and Ad Hoc networks [10]. For example, NIDS for a vehicular network [11].

Our objectives in this paper are to investigate several machine learning algorithms, investigate

deep learning network models, compare the different used algorithms and summarize the new challenges and issues in NIDS. To get the best result we will compare the algorithms that used the mostly known dataset called KDD Cup'99.

This paper is organized as follows: the next section describes the methodology, section 3 introduces the NIDS, section 4 explores the ML research then the DL researches in section 5. Section 6 describes the issues and challenges deduced from our study. And the last section will conclude the paper.

2. METHODOLOGY

Many surveys published in recent years proved that machine learning algorithms and deep learning algorithms are the best choices to detect intrusion. This study concentrates on the articles that evaluate the approaches of ML and DL over the dataset KDD Cup'99. We depend as a premier look over the statistics over Google Scholar. When we search the term "intrusion detection system machine learning" over Google Scholar, 13500 articles are given in this area of which 2600 are published in 2022. Therefore, the keywords in the search are as shown in table 1, and the number of articles is indicated for the selected keyword: machine learning or deep learning.

From this search, we noticed that the machine learning algorithm is used more than deep learning because as we know deep learning is considered a subset of machine learning. In addition, swarm intelligence is more used in machine learning for the feature selection process, in deep learning the feature selection is included in the algorithms. This does not also prohibit the authors to select the features as a preprocessing phase before applying the deep learning algorithm.

After these statistics, the selection of the articles used the KDD Cup'99 dataset to test the proposed approaches. Sections 4 and 5 explore the most recent articles. A search over Google Scholar shows that no survey was conducted on the intrusion detection system using machine learning or deep learning by focusing on KDD cup'99. But, when the search is done by using the keywords "intrusion detection system deep learning KDD cup'99 survey", we got 13600 articles, and when we used these keywords "intrusion detection system machine learning KDD cup 99 survey", we got 13100. This means that most of the research used the same dataset to test the proposed approaches.

In this report, we compare the most recent works in IDS using machine learning algorithms and deep learning algorithms over the dataset KDD Cup'99. The accuracy will be taken into consideration as a metric to compare these approaches.

Table 1: The search keywords on Google scholar.

Intrusion Detection System		
Keyword	Machine learning	Deep Learning
machine learning or deep learning	208,000	119,000
survey	376,000	89,200
machine learning swarm intelligence	33,800	26,800
feature selection	176,000	102,00

2.1 AI METHODS FOR NIDS

This section describes the general methodology which is common for machine and deep learning that is followed when designing the NIDS.
NIDS designed using machine learning or deep learning consists of general phases, these phases are shown in figure 1:

- Data preprocessing phase: in this phase, the data is transformed into a format suitable to be used by the selected algorithm. In general, the operations that may perform over the data are encoding and normalization. Some data need to be cleaned by removing unwanted data or duplicate data.
- Training phase: the algorithm of machine learning or deep learning is trained over 80% of the original data.
- Testing phase: the rest of the data which is 20% is tested by the algorithm to check the accuracy.

In addition, the following subsection will describe the commonly used dataset, the feature selection, and the evaluation metrics, to give the overall look over the NIDS.

2.2 Features Selection

KDD Cup'99 has 41 features, and not all features are important for building IDS, so we chose the most suitable one, by defining a subset of these features to achieve a high detection rate and low false alarms. The feature selection process is very important

process to creat the IDS, this process is essential to reduce the number of features.

Once you have completed the feature selection process using the suggested method. The feature set is trained using Machine Learning Classifiers. The classifier aims to distinguish between normal and attacks classes [8][10].

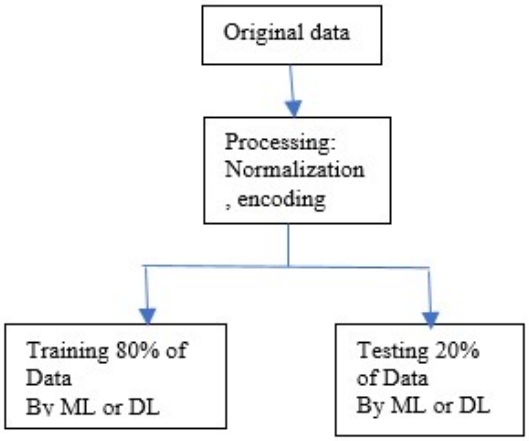

Figure 1: ML and DL Network-based IDS methodology

2.3 Evaluation Metrics

The evaluation of machine learning and deep learning is performed by several metrics. These metrics depend on different attributes arranged in an evaluation matrix as shown in figure 2:

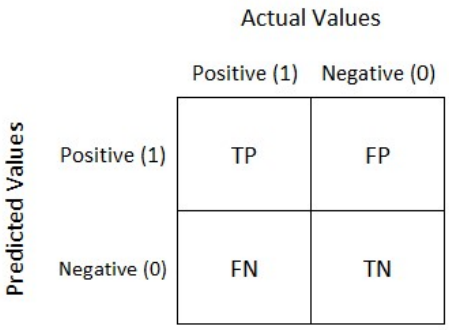

Figure 2: Evaluation matrix

These metrics include the actual and the predictive values:

1. True Positive (TP): the number of attacks that are correctly predicted as attack.
2. False Negative (FN): the number of attacks that are predicted as normal.
3. False Positive (FP): the number of normal traffics that are wrongly classified as an attack.
4. True Negative (TN): The number of normal traffics that are correctly classified as normal ones.

In addition to these evaluations, some common and very known metrics are Precision, recall, Accuracy, and F-measure. We are interested here in the accuracy which is equal to the sum of the true positive with the true negative divided by the summation of all values in the evaluation matrix.

2.4 Datasets

Several datasets are common and known to be used by the algorithms such as KDD Cup'99, Kyoto 2006+, NSL-KDD, UNSW-NB15, CIC-IDS2017, CSE-CIC-IDS2018. Our methodology is tested by the KDD Cup'99 dataset therefore we will explore the details of this dataset here.

KDD Cup'99 is the dataset used in the 3rd International Contest for Knowledge Discovery and Data Mining Tools, which was held in conjunction with the KDD-99 Fifth International Conference on Knowledge Discovery and Data Mining. This database contains a standard set of data to be audited, which includes a variety of simulated intrusions into a military network environment. Training and testing datasets are also available at UC Irvine KDD Archives [12][13].

KDD Cup'99 datasets were created by DARPA in 1999 using the network traffic recorded from the 1998 dataset. The complete dataset contains 41 features of a qualitative and quantitative nature, which contains 494020 numbers of instances (rows) of the dataset cases of nearly half a million (https://www.openml.org/d/1113). As for attack modes, 39 types of attacks are categorized into four main categories: 1) DOS: denial of service, e.g. synchronous flood. 2) R2L: Unauthorized access from a remote machine, for example, password guessing. 3) U2R: Unauthorized access to local super user (root) privileges.

3. NETWORK INTRUSION DETECTION SYSTEMS (NIDS)

A Network Intrusion Detection System (NIDS) is a tool that helps the network administrator to detect a harmful attack that may affect the network functions. NIDS monitors traffic passing through the network, analyzes this traffic to search for suspicious

activities and known threats, then sends alerts and stops them to protect the network from any attack.
NIDS could be classified according to the deployment method: the Host-based IDS which allows the IDS to be deployed on each host on the network and monitor the traffic arriving at this host. This deployment may degrade the performance and because it is executed on each host it may be more efficient. The other deployment method is the Network-based IDS which monitors the traffic arriving at the network to protect the entire network.

The detection method of intrusion may be classified into two categories: 1) signature-based detection system, which is also known as knowledge-based IDS because its idea is to store patterns of the attacks that may be generated, then to compare the arrived traffic with these patterns, if match then an intrusion is detected. This method is very efficient in detecting known attacks but if the attack is new then it will fail to detect it because its pattern is not stored on the database. The other category of detection method is the Anomaly-based detection system, this method is more efficient in detecting new attacks because its idea is to define the normal traffic and any deviations from normal are considered as an attack.

4. MACHINE LEARNING ALGORITHMS

ML algorithms are algorithms that enable the machine to learn automatically from a large dataset using a defined model and extract useful information. The common ML algorithms that are used in NIDS such as K-Nearest Neighbors (KNN) Gaussian Naive Bayes (GNB) Logistic Regression (LR) Decision Tree (DT) Random Forest (RF) Adaptive Boosting Classifier (ABC).

Many researchers proposed different algorithms of Machine Learning in IDS to reduce the rates of false positives and to increase the accuracy of an intrusion detection system.
The authors in [14] Proposed a model that evaluates several machine learning models (SVM, RF, Neural Network, XGBoost)on the KDD Cup'99. In addition, the model evaluated over ce the false alarm rate, and it proved its ability to reduce it. The feature selection is made by RF feature selection which gave high accuracy of 98.99.

The authors in [15] built two models for classifying the intrusion in the network. the first model uses the PSO and Decision Tree (DT), and the second uses the PSO and the K-Nearest Neighbor (KNN). the two models are tested and evaluated over the KDD Cup'99 datasets. the accuracy of PSO-KNN was 96.2, and the accuracy of POS-DT was 98.6. the PSO-KNN achieved the lowest score of 0.004 for the false positive rate.

The IDS proposed in [16], is tested over KDD Cup'99. it used similarity-based learning with the Siasame network. This network has two layers, it uses the constructive loss function. the model achieved an accuracy of 91%.In [17] researchers have proposed a cover feature selection algorithm for IDS, using a pigeon-inspired optimizer to take advantage of the selection process. Researchers have proposed a new method for continuous-optimizer coding and compared it to the traditional binary continuous intelligent swarm algorithms method. They used three standard data sets to evaluate the proposed algorithm: KDD Cup'99, NLS-KDD, and UNSW-NB15. The proposed algorithm showed massive superiority over many feature selection algorithms in terms of TPR, FPR, accuracy, and F-score. Table 2 summarizes the studies using machine learning with the obtained accuracy.
Most research using ML used optimization algorithm for feature selection. The authers in [18] used BAT algorithm with several ML algorithms. They test the algorithnm using DL and they get better results than ML. In [19], Harris Hawks optimization is proposed for feature selection and the tested model gave more robust results in term of accuracy, precesion and recall.

5. DEEP LEARNING METHODS

Deep learning methods are considered as a class of machine learning methods, they employs multi layers of information-processing stages in hierarchical manners for pattern classification and feature or representation learning instead of two layers in machine learning methods.

Deep learning consists of various networks, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), each having a different purpose and use [25]. Recently, deep learning techniques have been successfully applied in sevela domains such as text, audio, and visual processing [26] as well as contexts such as sentiment analysis [27], social network analysis, recommender systems [28], natural language processing, wireless networking [29], and so on. In addition, the DL

DL is a good tool used in IDS, where several researchers used it to reveal misuse detection. They use the DL to deal with most types of attacks and intrusions. The reviewed article [30]-[34] deal with the intrusion detection and anomaly detection.

Deep learning methods aim to classify the labeled data by applying the learned features, which are represented by a large amount of unlabeled data. When we talk about a large number of features, we are talking about deep learning for its ability to deal with it with reasonable results for its and l with unstructured data. Deep learning is faster when compared with Machine learning which takes a long time in learning and classifies the datasets of testing to deal with big data. In addition, that deep learning can learn directly from row data.

Deep learning is like machine learning teaches computers, to simulate the human brain and decide to do actually as the human is supposed to do. It is used to solve complex problems. The importance andDL's importance and use blem is its ability to deal with unstructured data.

Deep learning models consist of two categories of deep networks:

1- supervised learning models which consist of the following:

- Deep Neural Networks (DNNs): the processing of data is done in a complex way by passing it over many layers. The data modeling and the transfer of data from one level to another are done in sophisticated modeling.
- Convolutional Neural Networks (CNNs): this type is mostly used in IDS because it can learn from the data without feature extraction.
- Recurrent Neural Networks (RNNs): it is used to recognize the sequential characteristics of the data and use patterns to predict the next likely scenario. It is used in speech recognition and natural language processing.

2- unsupervised learning models

- Autoencoders: this type learns effective data encoding by training the network to ignore the noise. It may use in the generation of image data, image compression, and image processing.
- Restricted Boltzmann Machines (RBMs): is an algorithm useful for dimensionality reduction, classification, regression, collaborative filtering, feature learning, and topic modeling.
- Generative Adversarial modelling (GANs): in this model two neural network algorithms compete to get more accurate results.

The learned feature is modeled from the original data in Deep Learning approaches. No manyual feauture extraction will be needed, thus the DL can excute what is known from end to end manner. The most important things in DL is the network artchitecture, the parameter selection, and the optimization model. A comparison of various deep learning models is shown in Table 4.

The authors of the work in [4] built a prototype of merging an expert system with a neural network system, the expert system is for handli; thel available data and the neural network to increase accuracy, then the model has tested by installing it onand a real network. And another researcher [5] designed a system using also deep learning with 10,000 epochs salso usingom 9,462 records on the dataset and randomly selected 1000 records from their dataset for testing their system, their whole implementation of testingapproachraininfutook26.13 hours to complete,; thetesting results obtain 0.975569 accuracy.

In [29], the authors introduced DL-MAFID, which uses the multi-agent system and auto-encoder for dimension reduction to solve muti-class attack detection. they recognize five classes of attacks by using multilayer perceptron and K-Nearest Neighbor algorithm. the presented model was tested and the experiments achieved 99.95 accuracy with good detection time.

The authors in [30], proposed a deep learning model to extract the features that deal with the incomplete feedback in IDS, this allows us to decide the probable attacks to detect the intrusion. The authors test their model over the KDD Cup'99 dataset. They deal with intrusion in the cloud environment. Other algorithms are summarized in table 3.

6. ISSUES AND FUTURE CHALLENGES

This section explores the main issues and challenges in the ML/DL IDS:

1. Efficiency: the main criterion for any IDS is accuracy, this may be achieved by using a balanced dataset and up-to-date definition of the attacks. The NIDS must have a mechanism that can learn new features.
2. Real-time detection: the NIDS must be executed in real-time because the building model uses a very huge dataset the processing must be quick. To address this problem, the NIDS must be executed in high-performance computers with large storage.
3. Using of DL: from our study, the DL outperformed the ML and it still needs deep reinforcement learning.

This study explores the research that used KDD Cup'99. We noticed that the researcher still used this dataset to test their models, and many articles published after 2020 used this dataset. This dataset may be not suitable and must not be used further to test any model. The datasets gathered after are more powerful and efficient.

The dataset NSL-KDD is an improved version of KDD Cup'99 and cleaned to be more suitable, it contains more attacks but it is smaller than KDD Cup'99. In addition, it is a class imbalance between intrusion types.

The CSE-CIC-IDS2018 is a huge dataset that contains more attacks, it is a cleaned and class balance dataset that is more suitable to test any model than KDD Cup'99.

For the used approaches, the deep learning may still be needed to reinforcement learning and we advise to use the deep forest which more flexible than Deep Neural Network.

7. CONCLUSION

Cyber security is now paramount with the proliferation of data everywhere and in a massive amount over the storage spread all over the world. Data intrusion is the objective of hackers and intruders at all times. This paper introduces network intrusion detection systems, compares the different techniques, and points to the directions that may help choose the best algorithm. From this study, we conclude that machine learning algorithms always need an additional process for feature selection. Without this process, the algorithm may give wrong results. Machine learning consists of one layer to process the filtered data. While deep learning algorithms consist of several layers, the feature selection process is included in the deep learning algorithms; therefore, no need for any algorithm for feature selection, for this reason, most deep learning algorithms outperform machine learning algorithms.

The literature review shows that the trend is to use the DL tools to build the NIDS. The DL requires more computations and GPU increases the use of DL.

From the taken literature, we notice that the accuracy of the deep learning methods is always better than the machine learning methods and from our readings, the runtime complexity for deep learning is always better than the machine learning algorithms but it always needs to be more efficient for real-time detection.

REFERENCES:

[1] M. Masdari and M. Jalali, ''A survey and taxonomy of DoS attacks in cloud computing,'' Secur. Commun. Netw., vol. 9, no. 16, pp. 3724–3751, Nov. 2016.

[2] S. T. Zargar, J. Joshi, and D. Tipper, ''A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks,'' IEEE Commun. Surveys Tuts., vol. 15, no. 4, pp. 2046–2069, Mar. 2013.

[3] M. Almseidin, M. Alzubi, S. Kovacs, and M. Alkasassbeh, "Evaluation of Machine Learning Algorithms for Intrusion Detection System," no. Iv.

[4] H. Debar, M. Becker, and D. Siboni, "A neural network component for an intrusion detection system," Proc. Symp. Secur. Priv., pp. 240–250, 1992.

[5] J. D. Cannady, "Artificial neural networks for misuse detection," Proc. 21st Natl. Inf. Syst. Secur. Conf., pp. 368–381, 1998.

[6] M. Premkumar and T. V. P. Sundararajan, ''DLDM: Deep learning-based defense mechanism for denial of service attacks in wireless sensor net- works,'' Microprocessors Microsyst., vol. 79, Nov. 2020, Art. no. 103278.

[7] T. Jafarian, M. Masdari, A. Ghaffari, and K. Majidzadeh, ''A survey and classification of the security anomaly detection mechanisms in software defined networks,'' Cluster Comput., vol. 24, pp. 1–19, Jun. 2020.

[8] S. Iqbal, M. L. M. Kiah, B. Dhaghighi, M. Hussain, S. Khan, M. K. Khan, and K.-K. R. Choo, ''On cloud security attacks: A taxonomy and intru- sion detection and prevention as a service,'' J. Netw. Comput. Appl., vol. 74, pp. 98–120, Oct. 2016.

[9] R. V. K, R. Vinayakumar, S. Kp, and P. Poornachandran, "Evaluating Shallow and

Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security," 2018 9th Int. Conf. Comput. Commun. Netw. Technol., pp. 1–6, 2018.

[10] R. V. K, R. Vinayakumar, S. Kp, and P. Poornachandran, "Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security," 2018 9th Int. Conf. Comput. Commun. Netw. Technol., pp. 1–6, 2018.

[11] M. J. Kang and J. W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," PLoS One, vol. 11, no. 6, pp. 1–17, 2016.

[12] Hettich, S. and Bay, S. D. (1999). The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Department of Information and Computer Science

[13] Serinelli, B. M., Collen, A., & Nijdam, N. A. (2020). Training guidance with kdd cup 1999 and nsl-kdd data sets of anidinr: Anomaly-based network intrusion detection system. Procedia Computer Science, 175, 560-565.

[14] Priyavengatesh, M., & Kannan, R. (2022). An Efficient Intrusion Detection System Using Machine Learning Model. JOURNAL OF ALGEBRAIC STATISTICS, 13(3), 4984-5002.

[15] Aburomman, A. A., & Reaz, M. B. I. (2016). A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Applied Soft Computing*, *38*, 360-372.

[16] Kumar, G. R., Mangathayaru, N., & Narsimha, G. (2016). A novel similarity measure for intrusion detection using gaussian function. *arXiv preprint arXiv:1604.07510*.

[17] Eesa, A. S., Orman, Z. and Abdulazeez, A. M. (2015) 'A new feature selection model based on ID3 and bees algorithm for intrusion detection system', (April 2018). doi: 10.3906/elk-1302-53.

[18] Shen, Y., Zheng, K., Wu, C., Zhang, M., Niu, X., & Yang, Y. (2018). An ensemble method based on selection using bat algorithm for intrusion detection. The Computer Journal, 61(4), 526-538.

[19] Zivkovic, M., Bacanin, N., Arandjelovic, J., Rakic, A., Strumberger, I., Venkatachalam, K., & Joseph, P. M. (2022). Novel Harris Hawks Optimization and Deep Neural Network Approach for Intrusion Detection. In *Proceedings of International Joint Conference on Advances in Computational Intelligence* (pp. 239-250). Springer, Singapore.

[20] Chen, F., Ye, Z., Wang, C., Yan, L., & Wang, R. (2018, September). A feature selection approach for network intrusion detection based on tree-seed algorithm and k-nearest neighbor. In 2018 IEEE 4th international symposium on wireless systems within the international conferences on intelligent data acquisition and advanced computing systems (IDAACS-SWS) (pp. 68-72). IEEE.

[21] Ambusaidi, M. A. et al. (2016) 'Building an intrusion detection system using a filter-based feature selection algorithm', 9340(NOVEMBER 2014), pp. 1–13. doi: 10.1109/TC.2016.2519914.

[22] Manzoor, I. and Kumar, N. (2017) 'A feature re duce d intrusion detection system using ANN classifier', Expert Systems With Applications, 88, pp. 249–257. doi: 10.1016/j.eswa.2017.07.005.

[23] Selvakumar, B. and Muneeswaran, K. (2019) 'Firefly algorithm based feature selection for network intrusion detection', Computers and Security, 81, pp. 148–155. doi: 10.1016/j.cose.2018.11.005.

[24] Li, D. et al. (2019) 'International Journal of Information Management IoT data feature extraction and intrusion detection system for smart cities based on deep migration learning', International Journal of Information Management, 49(May), pp. 533–545. doi: 10.1016/j.ijinfomgt.2019.04.006.

[25] Selvakumar, B. and Muneeswaran, K. (2019) 'Firefly algorithm based feature selection for network intrusion detection', Computers and Security, 81, pp. 148–155. doi: 10.1016/j.cose.2018.11.005.

[26] Salih, A. A. (2019) 'Combining Best Features Selection Using Three Classifiers in Intrusion Detection System', 2019 International Conference on Advanced Science and Engineering (ICOASE), pp. 94–99.

[27] Kavitha, G. and Elango, N. M. (2020) 'An approach to feature selection in intrusion detection systems using machine learning algorithms', International Journal of e-Collaboration, 16(4), pp. 48–58. doi: 10.4018/IJeC.2020100104.

[28] Alazzam, H., Sharieh, A. and Sabri, K. E. (2020) 'A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer', 148. doi: 10.1016/j.eswa.2020.113249.

[29] F. Louati and F. B. Ktata, ‘‘A deep learning-based multi-agent system for intrusion detection,’’ Social Netw. Appl. Sci., vol. 2, no. 4, pp. 1–13, Apr. 2020

[30] A. Abusitta, M. Bellaiche, M. Dagenais, and T. Halabi, ‘‘A deep learning approach for proactive multi-cloud cooperative intrusion detection sys- tem,’’ Future Gener. Comput. Syst., vol. 98, pp. 308–318, Sep. 2019.

[31] S. Z. Lin, Y. Shi, and Z. Xue, ‘‘Character-level intrusion detection based on convolutional neural networks,’’ in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.

[32] Ravi, V., Chaganti, R., & Alazab, M. (2022). Recurrent deep learning-based feature fusion ensemble meta-classifier approach for intelligent network intrusion detection system. Computers and Electrical Engineering, 102, 108156.

[33] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. Ieee Access, 7, 41525-41550.

[34] Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. applied sciences, 9(20), 4396.

[35] Xu, C., Shen, J., Du, X., & Zhang, F. (2018). An intrusion detection system using a deep neural network with gated recurrent units. IEEE Access, 6, 48697-48707.

[36] J. Kim, N. Shin, S. Y. Jo, and S. H. Kim, ‘‘Method of intrusion detection using deep neural network,’’ in Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp), Feb. 2017, pp. 313–316.

Table2:Machine Learning Studies And Their Methods With The Obtained Accuracy

Reference	Feature Selection Method	The Used Algorithm	Accuracy
[19]	Bees algorithm (BA)	The interactive dichotomizer 3 (ID3) and Bees algorithm (BA)	92.002%
[20]	Mutual Information (MI)	Least Square Support Vector Machine based IDS (LSSVM-IDS)	99.79%, 99.91% 99.77%
[21]	information gain(IG) and correlation(CR)	Neural Network	98.79%
[22]	IWD algorithm	combination of IWD and SVM,	99.108%
[23]	Tree-Seed Algorithm(TSA)	Tree-Seed Algorithm and K-Nearest Neighbor ((KNN-TSA)	95.06%
[24]	Feature transfer method	Deep migration learning (RFID tag detection technology)	Normal 95.21% DOS 87.19% U2R 79.52 % R2L 80.83 % Probe 83.20%
[25]	Mutual Information (MI) and wrapper with Bayesian network,	Filter and wrapper algorithm with firefly algorithm	DoS 83.5% Probe 24.7% R2L 11.5% U2R 83.4%
[26]	Combine of Information Gain, Gain Ratio, and Correlation	KNN, NB, MLP	KNN 98.9% NB 93.3% MLP 96.5%
[27]	SVM	Guided Regularized Random Forest (GRRF)	98.55%
[28]	Sigmoid PIO and Modified binary Cosine PIO	Pigeon-inspired optimizer(PIO)	KDDCUP 99 96.0%

Table 3: Deep Learning Studies And Their Methods With The Obtained Accuracy

Reference	Feature Selection Method	The Used Algorithm	Accuracy %
[31]	-	CNN	85.07
[32]	-	RNN, GRU, LSTM, KPCA, random forest, SVM	99
[33]	-	DNN	95
[34]	-	DNN	
[35]	-	RNN-LSTM RNN-GRU	99.4 99.8
[36]	-	DNN	99.01

Table 4: Comparison Of Various Deep Learning Models [34]

Algorithm	Suitable data types	Supervised or unsupervised	Function
Autoencoder	Raw data, Feature vectors	Unsupervised	Feature extraction, feature reduction, denoising
BBM	Feature vectors	Unsupervised	Feature extraction, feature reduction, denoising
DBN	Feature vectors	Supervised	feature reduction, classification
DNN	Feature vectors	Supervised	Feature extraction, classification
CNN	Raw data, Feature vectors, matrices	Supervised	Feature extraction, classification
RNN	Raw data, Feature vectors, sequential data	Supervised	Feature extraction, classification
GAN	Raw data, Feature vectors	Unsupervised	Data augmentation, adversarial training.