Research Article

Classification of Cyber-Attack using Adaboost Regression Classifier and Securing the Network

Bdah Mohammed Mubarak AlShahrani[a] and Mohammad Tabrez Quasim[b]

[a] Masters students, Faculty of Computing and Information Technology, University of Bisha, Bisha, Saudi Arabia

[b]Faculty of Computing and Information Technology, University of Bisha, Bisha, Saudi Arabia

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: In recent years, with adverse development of technology leads to several security breaches. To withstand those security threats and breaches especially for cyber attacks resources are optimized with improved network lifetime. Those security challenges are lead to confidentiality, privacy, integrity and availability. To prevent cyberattacks artificial intelligence-based technology is evolved. To adopt appropriate cybersecurity wireless communication systems are intended to withstand threats and challenges. This paper, presented a deep learning-based classification technique for cyber attack detection. Deep learning structure involved in attack detection with proposed AdaBoost Regression Classifier (ABRC). The proposed ABRC with deep learning involved in estimation of attacks in the network security with deep learning structure. The proposed classifier model is involved in estimation of threats. The developed algorithm integrates AdaBoost and Regression classifier for threat detection and classification. The performance analysis expressed that proposed ABRC exhibits significant performance for cyber-attack detection than the existing deep learning technique.

Keywords: Cyberattack, Deep Learning, AdaBoost Regression Classifier, CICIDS, threats detection

1. Introduction

In the time of cybercrime damage losses by 2021, cyberattack is a critical issue, with a rise from 57.4 days to 93.2 days in the last two years [1]. Intrusion has averaged 6 trillion dollars a year for intrusion survey. The smart cities world market in 2017 is worth $40.1 trillion and their growth is projected to come to $97.9 billion by 2026[2], based on a study by Navigant Research. Most network consumers benefit from a variety of factors from the continued growth and widespread usage of the Internet. In the meantime, the widespread use of networks[3] makes network protection much more critical.

Security of the network is directly connected to computers, networks, systems, different data etc. in defense of unwanted entry and alteration. But in Finance, Electronic commerce and military, the increasing number of internet-connected systems make them targets for network attacks which lead to great risk and damage[4]. Essentially, powerful techniques to track and protect attacks and sustain network protection need to be provided. Also, it is normally important to process various kinds of attacks in different ways[5]. The key challenge in the field of network security, especially those unprecedented attacks, thus becomes how to distinguish various types of network attacks [6].

In recent years, researchers have used different types of ML approaches, without understanding their comprehensive characteristics, for classifying network attacks [7]. However, owing to their shortcomings in model complexity, conventional machine learning approaches are not capable of supplying distinctive attribute descriptors to characterize the issue of attack detection[8]. Through simulating the human brain with the layout of neural networks, which are called deep learning approaches for their general deep layer architecture to solve complicated problems, ML made a great breakthrough[9]. Google's AlphaGo is among these popular applications one of the strongest experiments for "go" play, which requires the intensity and strength of a standard deep learning structure, i.e. convolutionary neural networks[10].

Since deep learning is complicated in its original architectures and domain-oriented implementations, this paper is written to illustrate this to those who use deep learning approaches to research in the field of network security[11]. Using deep learning techniques, there is essentially a quantity of prior work focusing on attack detection. Including several analyses of literature, deeper learning on attack detection has been done. This paper, proposed a AdaBoost Regression classifier for cyber attack detection and prevention. The proposed algorithm involved in integration of both classifiers for improving accuracy and cyber attack prevention. The simulation analysis expressed that proposed ABRC exhibits improved cyber attack prevention mechanism compared with existing deep learning techniques.

2. Related Works

Cyber attack identification has been covered in detailed literature. The Bagged Tree and Gentle Boost have been compared with Ensemble machine learning methods for unbalanced data sets and are higher precisions and ROC values than other tree-related values [12]. [13] shows that the PCA is highly effective in distinguishing against KDD-99 and NSL-KDD data sets against cyber threats and regular Internet queries. The experimental findings also revealed the LDA error, which specifically has a weak covariance matrix estimation. In [14], accumulated entropy detection to filter Denial of Service attacks that are a big communication system issue. A high detection accuracy and a low false positive rate were displayed in the results. The findings have increased efficiency using the entropy field in the packet header via other detection approaches.

In [15], the Support Vector Machine (SVM) solution has been applied, which is a master training system (ML) which can supplement and reduce the efficiency of intrusion detection systems. A new technique for creating new identification rules was introduced in[16], which could distinguish both typical and unusual forms of attacks. Using the KDD-99 and NSL-KDD datasets, the efficiency of the DFEL system outperforms other traditional ML methods, however the UNSW-NB15 dataset, which is more dynamic and represents current internet traffic, still needs to be enhanced [17]. In [18] a range of reading materials are presented that explain the fundamental knowledge and history of the development of profound knowledge methods and their subsequent attack detection applications.

In [19] the focus is on illustrating strategies for attack prevention, malware analysis and spam detection related to intrusion detection. In their work[20], the study primarily in-depth learning approaches to ensure that the Internet of Thing technology has a good picture of different forms of cyber threats and their respective methods of identification. In [21], the study status of an intrusion detective system was subsequently evaluated and analyzed in four key datasets focused on deep learning techniques. They also study the related publications using the keywords "deep knowledge," "invasion" and "attack" in a systemic literature review that provides researchers with a large array of data history. Data is critical for the detection of intrusion [22]. In [23], 35 renowned network datasets were also identified and categorized into seven categories. They implement seven present models for each cypress where the accuracy and fake alarm rate based on the CSE-CIC-IDS2018 and the Bot-IoT are measured and compared. All the review papers, such as security programs, datasets, and databases, actually have their focus. In comparison to former approaches, our paper is focused on deep learning models and thus focuses on the methods of detecting attacks based on various deep learning architectures. Furthermore, we have a rational analogy and our specific review of the results of benchmark-based methods. In [24] applied the fuzzy clustering technique to generate different training subsets. As a result, training different neural networks on different training subsets could outperform some traditional machine learning algorithms such as decision tree and Naive Bayes.

3. Deep Learning Feature For Cyber Attack Classification

Deep learning consists of supervised and unsupervised learning methods and is based on several layers of artificial neural networks. The activation feature of each layer involves several neurons used for the processing of non-linear outputs. The approach is inspired by the brain's biological neuron structure but is closely connected to the patterns of information processing and coordination in the natural nervous system. Via the use of a hierarchical multi-level learning technique, deep learning algorithms extract significant abstract representations from raw data. The higher-level characteristics are more general and nuanced and are based on fewer abstract principles. The attributes in the lower stage of the learning hierarchy are representative. The complicated and high-level members are therefore essential as inputs for a regulated predictor. In figure 1 presented about overall architecture of deep learning

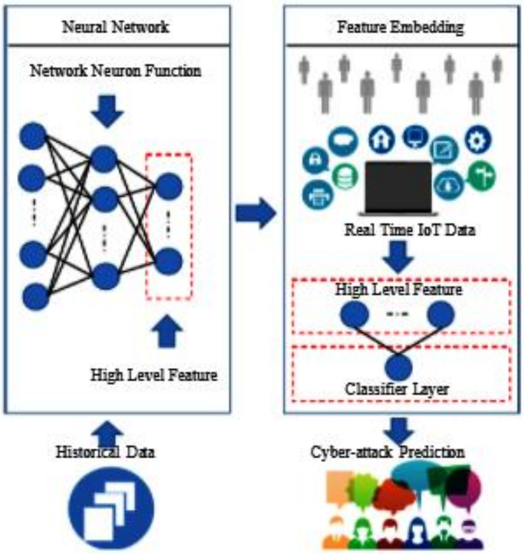

Figure 1: Architecture of Deep Feature Learning (DFL) for Attack Detection

The DFL approach was inspired both by the recent progress of transmission education in the area of visual categorization and word integration research. Figure 1 shows DFL architecture. The control value (from 0 to 1) must be calculated in III-A to match the training intensity and precision of prediction. Training data for the deep neural network are described. In the pre-trained profound learning network, embedding functions for small data sets with similar distributions have been developed. By using the high-level representative functions, however, limited data sets can take advantage of large volumes of data, as this can allow conventional master training classifiers to minimize their prediction time and improve detection accuracy.

3.1 Dataset for CyberAttack

This paper aimed to evaluate the attacks presented in the IDS system CICIDS 2019. Those are similar to UNSW-NB 15, KDD cup 19, and CICIDS 2017 dataset. IDS attack identification all features are not required, due to increased time for processing this reduces efficiency and accuracy. To overcome this limitation pre-processing in IDS is performed to eliminate redundant data for the optimal subset. Further pre-processing eliminates irrelevant features of the dataset for the original dataset without interfering with computational cost and accuracy. For reduction of dimensionality feature selection is adopted in Intrusion detection for simplification and time taken for dataset training. In figure 2 overall architecture for proposed deep learning classifier is presented.

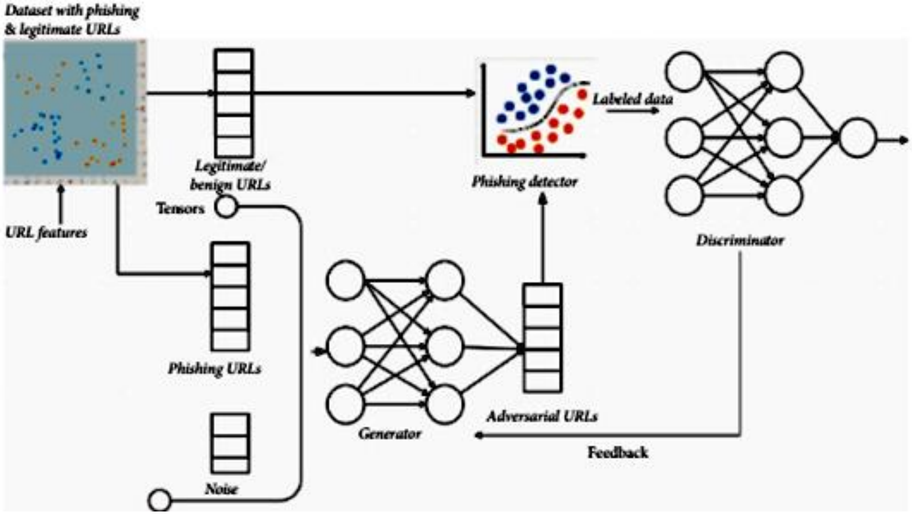

Figure 2: Overview of Deep Learning Architecture

The extraction of pre-processing is performed for collected data for attack extraction for minimal data dimensionality. Redundancy of irrelevant data involved in the reduction of feature selection. IDS features are selected based on three categories such as embedded model, filter, and wrapper. This research proposed a sigmoidal based ABRC approach for the learning process. The classification is performed with consideration of a few features such as training and testing dataset with the constraint of characteristics of data in terms of correlation effect, dependency, distance, and consistency. For identification of features of the cyber system utilizes the wrapping approach. In table 1 presented about distributed dataset for proposed ABRC.

Table 1: Dataset Distribution

	Data Distribution	**Count**
Training Set	Normal activity	67,343
	Anomaly	58,630
	DoS	45,927
	Probe Attack	11,656
	U2R (User to Root) Attack	52
	R2L (Root to Local)	995
Testing Set	Normal activity	9,710
	Anomaly	12,834
	DoS	7,458
	Probe Attack	2,422
	U2R (User to Root) Attack	67
	R2L (Root to Local)	2,887

Due to large number of data dimensionality model filter are applied. For computation of intensive factor proposed sigmoidal ABRC classifier involved in dataset processing by use of evaluation model for processing. For evaluation of relevant features data are integrated for minimal range. In figure 3 presented about deep learning mechanism for ABRC classifier performance.

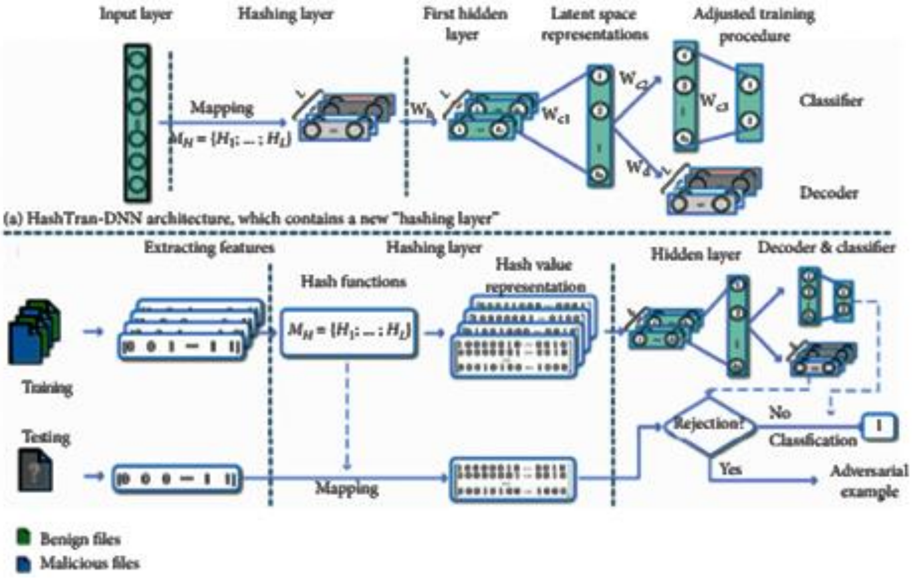

Figure 3: Deep Learning layer for ABRC

The configuration of constructed sigmoidal ABRC incorporates three different logs and data traffic those are all client logs, client configuration and attack logs. The selected CICIDS 2019 dataset consists of 14 attributes in which 12 attributes are utilized in this research those are presented in below table 2. The selected CICIDS 2019 dataset contains 172839 instances among those 153026 are used for analysis. The attributes for selected CICIDS2017 dataset were presented in below Table 2.

Table 2: CICIDS2019 dataset attribute Selection Features

Name of Attribute for CICIDS2019 dataset	**Description of attributes in CICIDS2019 dataset**
Src IP	IP address of the network address
Src Port	Port Source address
Dest IP	Destination network IP address
Dest Port	Port Destination
Proto	Protocol of Transport Layer
Date first seen	Data flow in-network for first time
Duration	Total flow duration
Bytes	Transmitted bytes count
Packets	Transmitted packets count
Flags	TCP flags concatenation
Class	Identification of class label whether normal, attacker and suspicious
Attack Type	Identification of attack type
Attack ID	Evaluation of attack id for class identification
Attack Description	Description of identified attack in the network

AdaBoost classifier performs effectively for identification of weak learner. Using training can identify strong classifier and using binary classification error are estimated. Through identification of weak learners AdaBoost improves the classification accuracy. This algorithm is based on the consideration of decision tree with consideration of various levels. The AdaBoost equation considered for this research is presented as follows in equation (1):

$$H = sign\left(\sum_{t} \alpha_t h_t\left(x_t\right)\right) \qquad (1)$$

Accuracy of classification is improved through integration of logistics regression with AdaBoost classifier. Generally, both classifiers perform binary and multiclass performance. Here, regression approach involved in prediction of logistic function, where those values are lies between 0 and 1, which means values below 0.5 are considered as 0. The general logistics equation considered are presented as follows in equation (2):

$$h_\theta(x) = g\left(\frac{1}{1+e^{-\theta T_x}}\right) \qquad (2)$$

This research intended to develop an efficient classification approach with identification and prediction of attacks in IDS. Here, logistics equation identified through regression model is utilized for deriving sigmoidal function incorporation with AdaBoost classifier with reduction of computational time in the process. Even though the developed sigmoidal based ABR approach is aimed to improve security of IDS, computational time also needs to be reduced with increased accuracy. Hence, this research for mathematical derivation chain rule and maximum likelihood property is integrated. The equation obtained after application of both properties are presented as follows in equation (3):

$$F'(x) = F'g(x)g'(x) \qquad (3)$$

Then simplified equation is presented as in equation (4),

$$P = P(k)(1-P(k)) \qquad (4)$$

For P maximum likelihood estimation is performed and presented as in equation (5) and (6),

$$\hat{l}(\theta;x) = \frac{1}{n}\sum_{i=1}^{n} \ln f(x_i|\theta) \qquad (5)$$

$$P = \sum \log P(k_i) + \sum \log(1-P_i(k_i)) \qquad (6)$$

For classifier negative terms will not be integrated hence removing negative terms on both sides of equation (7),

$$\sum P = \sum P_i \qquad (7)$$

Now, $P = (a_0 + a_1x_1 + a_2x_2 + ... + a_kx_k)\sum P_i$

4. Results And Discussion

The developed approach incorporates AdaBoost and regression classifier for conversion of attacks either 0's and 1's. Sigmoidal function improves the attack classification rate through which attack can be able to classify attacks in the network. With inclusion of sigmoidal approach CICIDS 2019 datasets are classified as training and testing dataset. Through classification dataset are involved for testing is provided as input for developed as classifier. Table 3 provides dataset for training and testing classification.

Table 3: Classification of training and Testing Data

Data Classes	Normal activity of network	Anomaly detection in the network	Identification of DoS	Evaluation of Probe	Evaluation of U2R (User to Root)	Evaluation of R2L (Root to Local)
Training set	66,586	51,2622	42,455	10,7655	46	856
Testing set	8,456	11,152	6,7655	2,756	61	1,657

In table 4.1 provides data classified for consideration of attack in the collected CICIDS 2019 dataset. The collected dataset consists of 4 attacks such as DoS, U2R, R2L and Probe attack. This implies that collected dataset includes both insider and outside attack, which in turns improves the proposed system performance. Based on the classification of attacks training data and testing data were implemented as shown in figure 4 and 5.

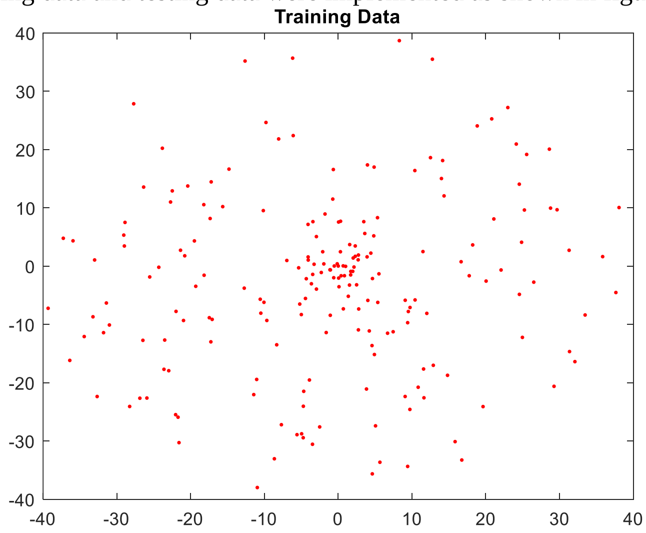

Figure 4: Training Set for CICIDS 2019

This research uses integrated classifier with sigmoidal function hence regression classifier for CICIDS 2019 includes classification of actual and predicted class. Based on the consideration of regression model AdaBoost approach is involved in classification of attack in the network. Figure 4.3 provides classification of testing data for selected dataset CICIDS 2019. In that figure it is observed that among the coverage range of 100 meters of WSN huge amount of training data was observed in 0 meters. Apart from the selected region minimal range of dataset was observed.

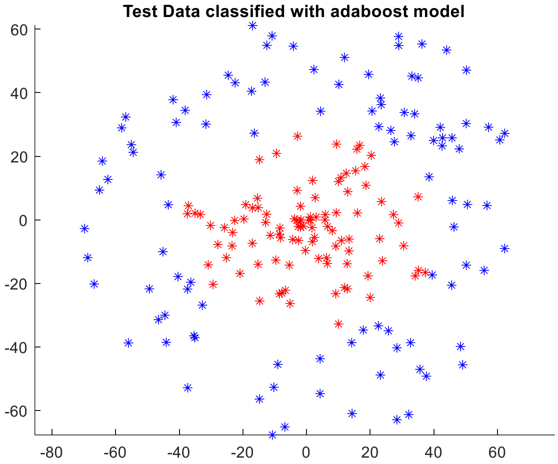

Figure 5: Testing Data set Classification

It is necessary to estimate the dataset utilized for testing of dataset for the proposed approach. The developed model uses CICIDS 2019 dataset for classification of attack dataset for identification of attacks in the network. The above figure 5 provides the graphical representation of classified testing data utilized for classification of attacks in the network. In figure 5, red dots illustrated the training data and blue dots indicate testing data. As shown in figure 4 training data were spread till 40 and testing data were observed till 60. In figure 6 presented about deep learning feature for error estimation.

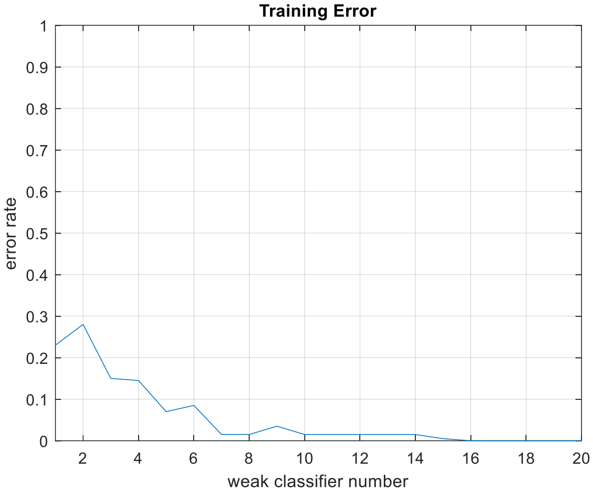

Figure 6: Error Estimation

After processing of dataset utilized CICIDS 2019 dataset error rate are estimated for classification of attack. The proposed approach estimate weak classifier with utilization of AdaBoost approach. For the presented dataset error rate is observed as 0.275 for the weak classifier number of 2. After that weak classifier count with minimal error rate of 0.1. Figure 7 estimate the training error value to minimal rate.

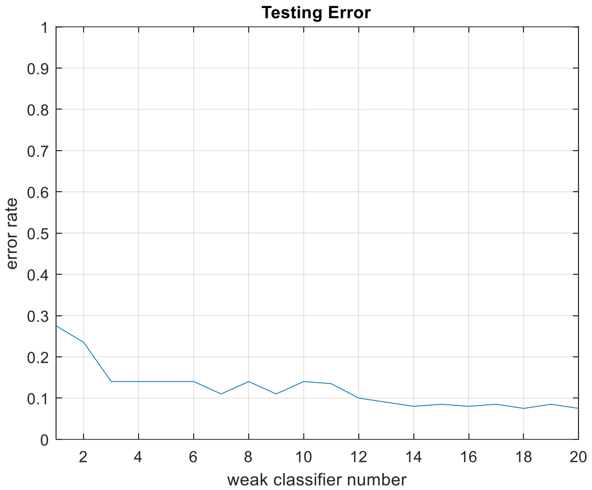

Figure 7: Testing Error estimation

The above figure 7 offers testing error rate observed for developed CICIDS 2019 dataset. For the weak classifier error rate is observed as 0.29 and for another classifier count it is reduced.

4.1 Evaluation Metrics

The main objective of this section is to test ABRC detection time and precision. The detection efficiency is also presented by authors in algorithm 1. The assessment criteria are important to determine the right method for identifying cyber attacks. TP refers to the correct detection of intrusion, and FP means that normal traffic is considered a cyber attack. TN is accurately defined as regular traffic and FN is a failed disclosure of intrusion. The findings of the tests use the following efficiency measures.

Accuracy: The metric assesses the percentage of correctly categorized internet traffic. It is a fraction that is divided by the total number of instances in the dataset.

Recall: This calculation represents the capacity of the classifier to detect cyber attacks, also known as sensitivity, which is critical.

Precision: This metric relates to the capacity of the classifier to unconditionally satisfy the normal request, also called specificity.

Processing Time Change: Time of detection is depends on training and testing time. Time shift (TC) is defined as the fraction of the detection time of the classifier without ABRC (T) minus the detection time of the classifier after ABRC, divided by the processing time without ABRC.

In table 4 presented about deep learning ABRC with existing classification model. The analysis is based on consideration of different performance metrics.

Table 4: Overall comparative analysis

Model	**Accuracy**	**Precision**	**Recall**	**Processing time (sec)**
Gradient Boosting	89.63	87.57	89.36	112
k-nearest neigbour	90.13	89.89	90.73	70
Decision Tree	91.48	91.86	91.94	80
Logistic Regression	91.45	93.56	90.78	63
Support Vector Machine	93.78	93.68	94.78	47
ABRC	95.87	95.93	96.74	33

In figure 8 overall comparative analysis of proposed ABRC classifier with existing classifier is presented.

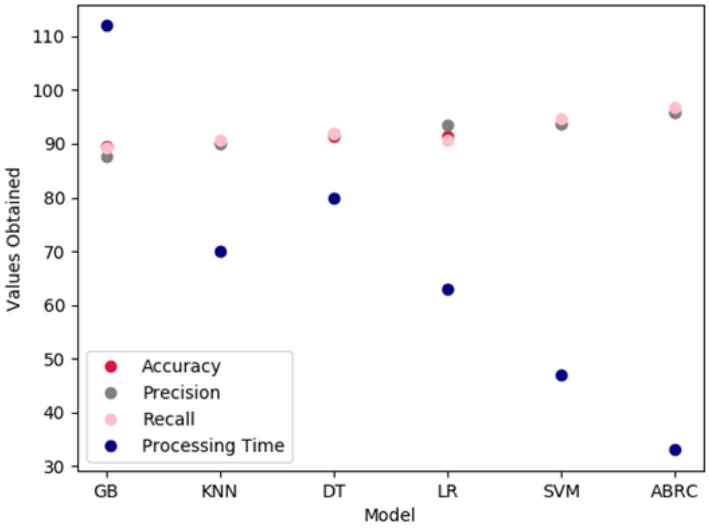

Figure 8: Comparative Analysis

5. Conclusion

This paper proposes an ABRC for real time cyber attack detection, a new deep learning approach. This approach's central concept is to map the initial low level to a high standard r(r < d). The consequence in our experiment presents the power of high precision and significant savings in time. With proper settings the IDS will adjust detection time and performance. This method may also be used in a situation where vast quantities of data are present and where the real time forecast is required. This approach will be implemented for actual equipment to deter cyber attacks in our future work. In reducing data dimensionalities, we will aim to boost the ABRC's ability.

References

1. AlDairi, A. (2017). Cyber security attacks on smart cities and associated mobile technologies. *Procedia Computer Science*, *109*, 1086-1091.
2. Han, M., Duan, Z., & Li, Y. (2017). Privacy issues for transportation cyber physical systems. In *Secure and Trustworthy Transportation Cyber-Physical Systems* (pp. 67-86). Springer, Singapore.
3. Idhammad, M., Afdel, K., &Belouch, M. (2017). Dos detection method based on artificial neural networks. *International Journal of Advanced Computer Science and Applications*, *8*(4), 465-471.
4. Aleesa, A. M., Zaidan, B. B., Zaidan, A. A., & Sahar, N. M. (2020). Review of intrusion detection systems based on deep learning techniques: coherent taxonomy, challenges, motivations, recommendations, substantial analysis and future directions. *Neural Computing and Applications*, *32*(14), 9827-9858.
5. Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018, May). On the effectiveness of machine and deep learning for cyber security. In *2018 10th international conference on cyber Conflict (CyCon)* (pp. 371-390). IEEE.
6. Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122.
7. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., &Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419.
8. Xu, X., He, C., Xu, Z., Qi, L., Wan, S., & Bhuiyan, M. Z. A. (2019). Joint optimization of offloading utility and privacy for edge computing enabled IoT. *IEEE Internet of Things Journal*, *7*(4), 2622-2629.
9. Xu, X., Liu, Q., Zhang, X., Zhang, J., Qi, L., & Dou, W. (2019). A blockchain-powered crowdsourcing method with privacy preservation in mobile environment. *IEEE Transactions on Computational Social Systems*, *6*(6), 1407-1419.
10. A. Sampathkumar, Mulerikkal, J. &Sivaram, M. Glowworm swarm optimization for effectual load balancing and routing strategies in wireless sensor networks. Wireless Networks, vol. 26, no. 6, 4227–4238 (2020). https://doi.org/10.1007/s11276-020-02336-w.
11. Xu, X., Liu, X., Xu, Z., Dai, F., Zhang, X., & Qi, L. (2019). Trust-oriented IoT service placement for smart cities in edge computing. *IEEE Internet of Things Journal*, *7*(5), 4084-4091.
12. Wang, C., Chen, Z., Shang, K., & Wu, H. (2019). Label-removed generative adversarial networks incorporating with K-Means. *Neurocomputing*, *361*, 126-136.
13. Meng, T., Wolter, K., Wu, H., & Wang, Q. (2018). A secure and cost-efficient offloading policy for Mobile Cloud Computing against timing attacks. *Pervasive and Mobile Computing*, *45*, 4-18.

14. Meng, T., Wolter, K., Wu, H., & Wang, Q. (2018). A secure and cost-efficient offloading policy for Mobile Cloud Computing against timing attacks. *Pervasive and Mobile Computing*, *45*, 4-18.
15. 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.
16. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... &Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*.
17. Farahnakian, F., &Heikkonen, J. (2018, February). A deep auto-encoder based approach for intrusion detection system. In *2018 20th International Conference on Advanced Communication Technology (ICACT)* (pp. 178-183). IEEE.
18. Sampathkumar*, Vivekanandan. P "Gene Selection Using PLOA Method In Microarray Data For Cancer Classification" Journal of Medical Imaging and Health Informatics 9, 1294-1300. Scopus Indexed, Impact Factor 0.549
19. Papamartzivanos, D., Mármol, F. G., &Kambourakis, G. (2019). Introducing deep learning self-adaptive misuse network intrusion detection systems. *IEEE Access*, *7*, 13546-13560.
20. Erpek, T., Sagduyu, Y. E., & Shi, Y. (2018). Deep learning for launching and mitigating wireless jamming attacks. *IEEE Transactions on Cognitive Communications and Networking*, *5*(1), 2-14.
21. Peng, W., Kong, X., Peng, G., Li, X., & Wang, Z. (2019, July). Network intrusion detection based on deep learning. In *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)* (pp. 431-435). IEEE.
22. Yang, H., & Wang, F. (2019). Wireless network intrusion detection based on improved convolutional neural network. *Ieee Access*, *7*, 64366-64374.
23. Tang, D., Tang, L., Shi, W., Zhan, S., & Yang, Q. (2020). MF-CNN: a New Approach for LDoS Attack Detection Based on Multi-feature Fusion and CNN. *Mobile Networks and Applications*, 1-18.
24. Zheng, Z., Tang, D., Wang, S., Wu, X., & Chen, J. (2020, August). An Efficient Detection Approach for LDoS Attack based on NCS-SVM Algorithm. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)* (pp. 1-9). IEEE.