Effective Intrusion Detection in IoT Environment: Deep Learning Approach

Sualihah Jan[1], Faheem Masoodi[2], Alwi M Bamhdi[3]

Department of Computer Science, University of Kashmir, India[1,2]
Department of Computer Science, Umm AL-Qura University, Saudi Arabia[3]
Corresponding author: Faheem Masoodi, Email: masoodifahim@uok.edu.in

IoT is among the most important technologies in recent times, which has evolved for the good of human beings. It reduces the human efforts in every domain possibly by connecting things, making people work and live smarter. The use of smart devices in a variety of applications has helped pave the way for pervasive computing, which offers huge human, economic, and other advantages. However, these advantages come with a number of drawbacks that must be addressed, one of which is security.The issue addressed in this research will be an effective Intrusion Detection System deployed in the Internet of Things environment. Though, various Intrusion detection systems are available, some are using learning methods. However, they face challenges such as a lack of relevant data, with some relying on the KDD dataset, which isn't designed specifically for IoT. This paper presents Deep Learning-based techniques, such as DNN and LSTM-RNN classifier, for detecting classes of assaults and anomalies in a simulated smart environment and classifying them as abnormal or normal using the BoT-IoT dataset. The proposal framework was created with the Google Colab platform. The experimental results for all the attack classes using DNN and LSTM-RNN classifiers have achieved 99.7% and 99.8% accuracies respectively.

Keywords: IoT, Google Colab platform, KDD dataset, DNN and LSTM-RNN classifiers.

2021. In Raju Pal & Praveen K. Shukla (eds.), *SCRS Conference Proceedings on Intelligent Systems*, 495–502. Computing & Intelligent Systems, SCRS, India.
https://doi.org/10.52458/978-93-91842-08-6-47

1 Introduction

Internet of Things is indeed a universal information skeleton based on Internet which makes it easier to trade products and services. The Internet of Things (IoT) provides an IT infrastructure that allows for the interchange of “things” between various connected devices, bridging the gap between physical objects and their representation in information systems [1]. The Internet of Things(IoT) is infiltrating as well as flourishing across every aspect of human lives, including homes, our education sector, transportation sector, and healthcare sector [2]. When Internet of Things (IoT) applications grow more common, security vulnerabilities in IoT raise since most of the connected devices are not build to withstand privacy and security, which has resulted in a slew of privacy and security issues in IoT system. We can say that Internet of Things has the potential to make human existence easier at the expense of security [3].

In an IoT network, data is produced by a variety of devices, analyzed in a variety of ways, shifted to different places, and acted on by applications. There are a variety of structures available that have been endorsed by a variety of professionals [4]. Some of the attacks occurring in IoT environment based on the CISCO IoT architecture levels is shown in the below Fig. 1. Further these attacks are divided into two categories: i) *Active attack*: The attacker disrupts the network's performance during an active attack by stealing data during communication e.g., Spoofing Attack, Denial of Services, Sinkhole, and Sybil attack. ii) *Passive attack*: The communication channels are watched, and the passive attacker steals information based on their usage history e.g., Eavesdropping, Brute force and traffic analysis [5].

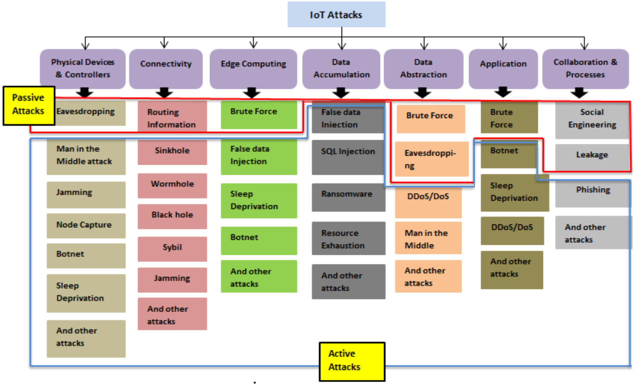


Fig. 1. IoT Attack Taxonomy

Malicious cyber attacks pose major security risks, necessitating the development of a new, adaptable, and reliable intrusion detection system (IDS) [6]. In real time, an Intrusion detection system (IDS) is the one which detects and categorizes intrusions, attacks, and violations of security protocols at the host and network level architecture [7]. On the basis of location of IDS, Intrusion detection is divided into two categories: i) *Network-based IDSs*: can analyze a vast traffic network with just a few well-placed devices or nodes, and have very minimal overhead.ii)*Host-based IDS*: is an IDS system that detects assaults by using system activity [8]. This is done by checking various log files running on the host machine. IDS are also classified on the basisof type of detection used: i) *Signature-based IDS*, additionally termed as "misuse intrusion detection" or "knowledge-based intrusion detection," is founded on idea where signatures are created for attack patterns [9]. To detect these attacks, signatures are kept securely inside a database, from which upcoming traffic data packets are later compared to the patterns present in the database. It benefits in increased detection rate ofknown attacks since signatures for those attacks are available in the created database [10]. This technique, on the other hand, seems unable to detect new attacks due to the lack of signature patterns. ii) *Anomaly-Based Detection* is a form of IDS that monitors network traffic for anomalies and recognizes data that is inaccurate, invalid, or otherwise unusual [11]. This method is handy for detecting undesirable traffic that isn't known to the user. iii) *Hybrid-based IDS* was created to address the shortcomings of Signature based IDS and Anomaly based IDS by combining these two to detect both undiscovered and recognized assaults [12]. The Machine learning techniques provide flexibility, adaptability, and low CPU load which will motivate us to develop a variety of analytical models for attack and anomaly detection that are more accurate and have lower false alarm rates [13]. One of the most powerful subcategories of Machine Learning is Deep Learning (DL). Deep Learning models the way human brain processes data to detect things, recognize speech, translate languages, and make judgments [14]. The hierarchical nature of deep learning, unlike other standard linear programs in machines, allows it to take a nonlinear approach to data processing, processing input over a number of layers, each of which will integrate following tiers of additional information [15]. Because of their hierarchical structure, deep networks can achieve more precision in terms of classifications and predictions. When used with IDS, DL networks may identify new attacks and anomalies with superhuman efficiency [16]. The technology's key benefit over machine learning is that it eliminates the need for manual feature selection and allows for the modeling of non-linear connections [17].

2 Literature Survey

Numerous research papers offered by various writers will aid in expanding our understanding of the various security aspects of IoT, as well as the ways required to address them. For unsupervised feature learning, a non-symmetric deep autoencoder (NDAE) was proposed by the authors [18]. In addition, they presented one of a kind deep learning categorization model built with layered NDAE. The suggested classifier was developed in TensorFlow with GPU support and tested on NSL-KDD and the KDD Cup '99 datasets.

Diro et al. [19] came up with a deep learning model capable of detecting distributed assaults on NSL-KDD dataset, which employs deep learning's self-teaching capabilities for identifying attacks at fog devices. In comparison to centralized systems, the results demonstrated that distributed attack detection provided superior accuracy.

Farhanullah et al. [20] proposed a deep neural network based on TensorFlow for detecting software's copyright theft as well as other malware-related assaults occurring in an industrialized IoT infrastructure. This combination solution included using Google Code Jam's dataset where in deep learning was used to detect duplication in program code, and deep CNN to detect harmful actions using colored image visualization. In an IoT network, traffic classification is critical for guaranteeing security.

Akbar et al. [21] resolved the imbalance of class problem for IDS in IoT, CSSAE was used for feature selection on KDDCup99 and NSL-KDD. CSSAE not only distinguished between benign and malicious classes, but also between distinct types of malicious classes present. CSSAE shows better performance in detecting attacks whose number is low such as Probe, U2R, R2L.

Bot-IoT, a novel dataset that comprises of both real and simulated IoT network activity, as well as numerous forms of assaults was presented by the authors [22]. They also proposed a realistic testbed environment for fixing the problems with the current dataset i.e., capturing entire network information, correct labeling, and recent plus complicated attack varieties. Finally, they compared the BoT-IoT dataset to the benchmark datasets to assess its forensics reliability using various statistical and machine learning methodologies.

3 Methodology

We will start with the dataset preprocessing followed by the second step, which is training and testing of two Deep learning classifiers DNN and LSTM-RNN on the dataset. Third and final step will be evaluation of the results on various parameters which will be discussed in the next section. The overall methodology is depicted in Fig. 2.

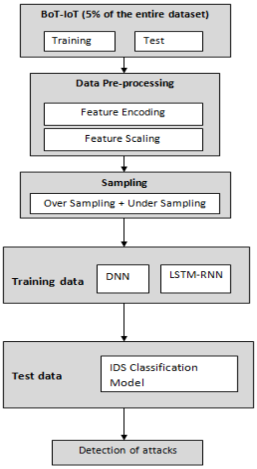
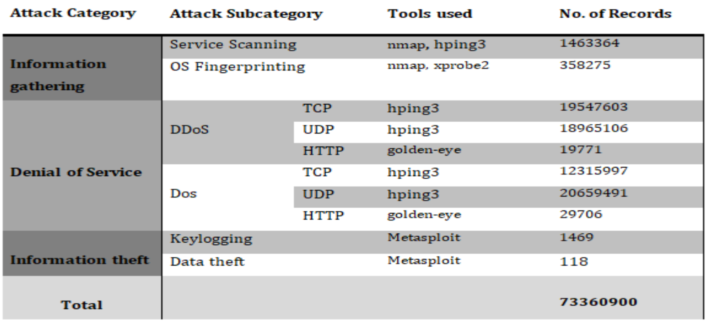


Fig. 2. Proposed IDS Model

A. Dataset and Data Pre-processing

In our study, we will use "10-best Training-Testing split" (5% of the entire dataset) which is one of the subsets of BoT-IoT dataset. It is made up of two CSV files: training and testing. Dataset for training contains 29,34,817 records and 19 features whereas Testing dataset contains 7,33,705 records and 19 features. Before we evaluate the DNN model on the "10-best Training-Testing split" dataset, we pre-process the data. Total attacks sum to 7,33,60,900 and normal records upto 9,543 [22][23]. Table 1 depicts attack traffic and associated distribution.

Table 1. BoT-IoT Traffic distribution



Attack Category	Attack Subcategory		Tools used	No. of Records
Information gathering	Service Scanning		nmap, hping3	1463364
	OS Fingerprinting		nmap, xprobe2	358275
Denial of Service	DDoS	TCP	hping3	19547603
		UDP	hping3	18965106
		HTTP	golden-eye	19771
	Dos	TCP	hping3	12315997
		UDP	hping3	20659491
		HTTP	golden-eye	29706
Information theft	Keylogging		Metasploit	1469
	Data theft		Metasploit	118
Total				73360900

Dataset preprocessing is an important step before feeding data to a machine of any sort. Actual data is transformed to machine-readable form. Pre-processing steps are performed to reduce computational load of the dataset, with increased efficiency of our model."10-best Training-Testing split" dataset was subjected to the following data preprocessing steps: i) Imputation of missing values. ii) Dropping the features that are not required. iii) Dropping the particular record from feature column with the least count. iv)Replacing port values with 0. v) Converting data types.vi)Feature Encoding i.e, converting categorical variables into numerical labels. vii) Feature Scaling is technique which brings all the feature values in one range. viii) Balancing the Dataset.

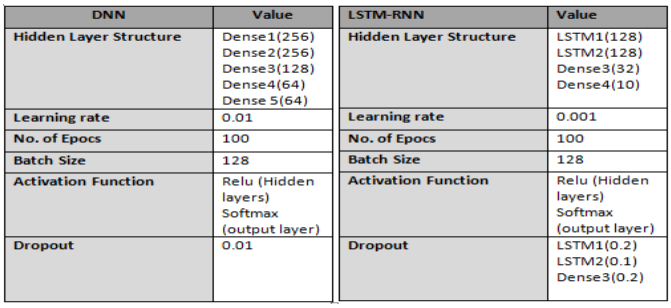
B. Classification

We used two Deep learning models DNN and LSTM-RNN for detecting multi-class intrusions in our dataset.

Deep Neural Network has evolved from simple artificial neural network and has increased number of hidden layers in between input and output layers (DNN). Each node is a neuron, followed with a nonlinear activation function, with the exception of the input nodes. Each new layer is made up of a set of nonlinear functions that are the weighted sum of all the previous layer's outputs (completely linked). Back propagation comes under a supervised learning technique which is used by MLP to train the network [24].

RNNs are a form of neural network that is both powerful and reliable, and they are among the most promising algorithms now in use because they are included with an internal memory. RNNs can be used to model sequence data. Architecture of RNN is same as of a DNN but here hidden units are capable of synthesizing the information from the previous hidden node to the present one. Hidden units can also be viewed as network attached storage, as they keep track of everything from start to finish [25].The usage of "LSTMs," a particularly specific type of recurrent neural network that performs far better than the regular version for many tasks, is critical to these accomplishments [26]. RNNs with the capability of learning long-term dependencies are known as Long Short Term Memory Networks (LSTMs). It also refines 'vanishing gradient' and 'exploding gradient problems' of basic RNN[27].

Table 2. Hyperparameters used in our work



DNN	Value	LSTM-RNN	Value

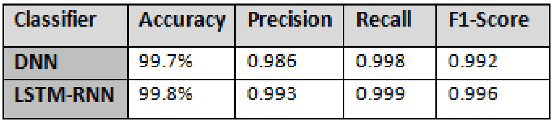
Hidden Layer Structure	Dense1(256)
	Dense2(256)
	Dense3(128)
	Dense4(64)
	Dense 5(64)
		Hidden Layer Structure	LSTM1(128)
			LSTM2(128)
			Dense3(32)
			Dense4(10)

Learning rate	0.01	Learning rate	0.001
No. of Epocs	100	No. of Epocs	100
Batch Size	128	Batch Size	128
Activation Function	Relu (Hidden layers) Softmax (output layer)	Activation Function	Relu (Hidden layers) Softmax (output layer)

Dropout	0.01	Dropout	LSTM1(0.2)
			LSTM2(0.1)
			Dense3(0.2)

4 Results

In our proposed work, we have taken "5% of the entire dataset" having 19 columns, to make the computing process of the BoT-IoT dataset easier. This 5% version of dataset consists of four files totaling around 3 million records from which we used already split Train and Test version. The platform used for carrying our experimentations was Google Colab. Deep Learning classifiers namely DNN and LSTM-RNN were trained and tested on 29,34,817 and 7,33,705 records of used dataset. The performance of models was evaluated on various parameters.



Classifier	Accuracy	Precision	Recall	F1-Score
DNN	99.7%	0.986	0.998	0.992
LSTM-RNN	99.8%	0.993	0.999	0.996

The graphical representation of results are shown in the below figures and charts.

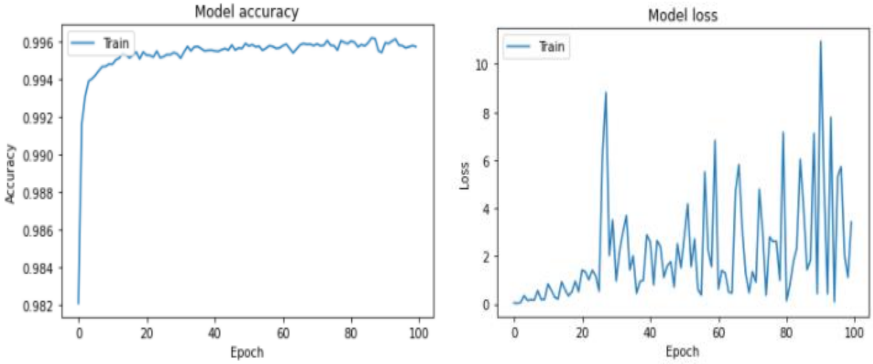


Fig. 3. Model accuracy and Model loss during training of DNN classifier

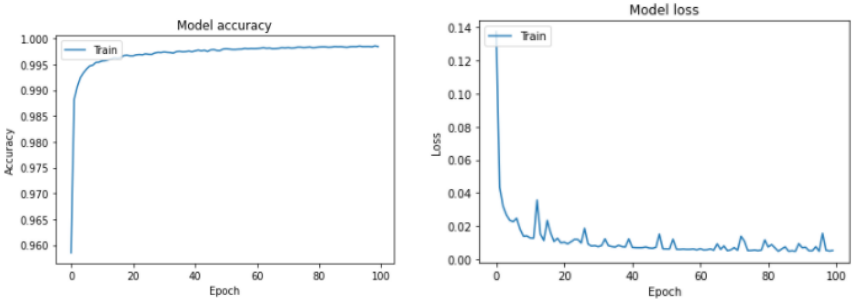


Fig. 4. Model accuracy and Model loss during training of LSTM-RNN classifier

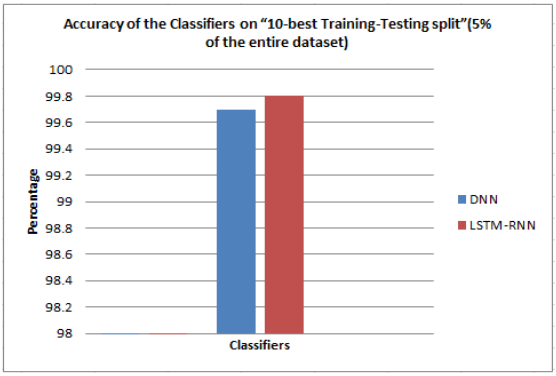


Fig. 5. Accuracy of Deep Learning classifiers

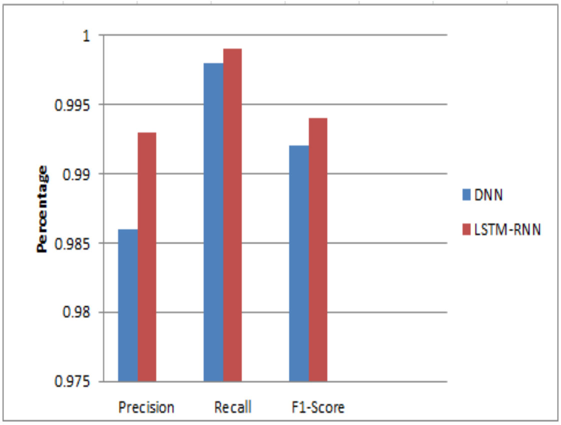


Fig. 6. Performance chart

5 Conclusion and Future Scope

In this research, we trained and tested two Deep Learning models viz. Deep Neural Network and Long Short Term Memory Recurrent Neural Network for intrusion detection. The learning methods of these two classifiers were compared on a 5% dataset version from BoT-IoT dataset, produced using statistical operations, and finally comprised of 10 best features. The performance evaluation was carried using four metrics which are accuracy, precision, recall and f1-Score on each of the Classifiers. Results obtained have shown that LSTM-RNN has outperformed DNN in terms of accuracy, precision, recall and f1-Score.

For Future, the available assessment mechanisms for hyper parameters will be examined in depth as well as worked on to increase the efficiency of our model.

References

[1] Weber, R. H. and Weber, R. (2010). Internet of Things. *Springer, Berlin, Heidelberg*.

[2] Ferrag, M. A. et al. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, 50:102419.

[3] Mishra, N. and Pandya, S. (2021). Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review. *IEEE Access*, 9:59353–59377.

[4] Bamhdi, A. M., Abrar, I. and Masoodi, F. (2021). An ensemble based approach for effective intrusion detection using majority voting. *Telkomnika,* 19(2):664–671.

[5] Bokhari, M. U. and Masoodi, F. (2012). BOKHARI: A new software oriented stream cipher: A proposal. In the *Proceeding of WICT,* 128–131.

[6] Idrissi, I., Azizi, M. and Moussaoui, O. (2020). IoT security with Deep Learning-based Intrusion Detection. Systems: A systematic literature review. In the *4th International Conference on Intelligent Computing in Data Sciences.*

[7] Bokhari, M. U. and Masoodi, F. (2012). BOKHARI: A new software oriented stream cipher: A proposal. In the *Proceeding of WICT,* 128–131.

[8] Masoodi, F., Alam, S. and Bokhari, M. U. (2011). SOBER Family of Stream Ciphers: A Review. *International Journal of Computer Applications*, 23(1):1–5.

[9] Abrar, I., Ayub, Z. and Masoodi, F. (2021). Current Trends and Future Scope for the Internet of Things. *In book: Internet of Things in Business Transformation: Developing an Engineering and Business Strategy for Industry 5.0, 185–209.*

[10] Masoodi, S. T. et al. (2019). Security and privacy threats, attacks and countermeasures in Internet of things. *International Journal of Network Security & Its Applications,* 11(2):67–77.

[11] Ahanger, A. S., Khan, S. M. and Masoodi, F. (2021). An Effective Intrusion Detection System using Supervised Machine Learning Techniques. In the *Proceeding of 5th ICCMC,* 1639–1644.

[12] Teli, A. T. and Masoodi. (2021). Security Concerns and Privacy Preservation in Blockchain based IoT Systems: Opportunities and Challenges. In the *ICICNIS*, 29–36.

[13] Khraisat, A. et al. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cyber Security*, 2:20.

[14] Pandow, B. A., Bamhdi, A. M. and Masoodi, F. (2020). Internet of Things: Financial Perspective and Associated Security Concerns. *International Journal of Computer Theory and Engineering*, 12(5):123–127.

[15] Masoodi, F. et al. (2021). Machine Learning for Classification analysis of Intrusion Detection on NSL-KDD Dataset. *Turkish Journal of Computer and Mathematics Education,* 12(10):2286–2293.

[16] Masoodi, F. S. and Bokhari, M. U. (2019). Symmetric Algorithms I. *Emerging Security Algorithms and Techniques*, 79–95.

[17] Teli, T. A. and Masoodi, F. (2021). Blockchain in Healthcare: Challenges and Opportunities. *SSRN Electronics Journal*, 1–6.

[18] Malhotra, P. et al. (2021). Internet of Things: Evolution, Concerns and Security Challenges. *Sensors (Basel)*, 21(5):1809.

[19] Shone N. et al. (2018). A Deep Learning Approach to Network Intrusion Detection. *IEEE Transactions on Emerging Topics in Computational Intelligence,* 2:41–50.

[20] Diro A. A. and Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems*, 82:761–768.

[21] Ullah F. et al. (2019). Cyber Security Threats Detection in Internet of Things Using Deep Learning Approach. *IEEE Access,* 7:124379–124389.

[22] Telikani, A. and Gandomi, A. H. (2019). Cost-sensitive stacked auto-encoders for intrusion detection in the Internet of Things. *Internet of Things*, 100122.

[23] Koroniotis, N. et al. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. *Future Generation Computer Systems*, 100:779–796.

[24] Gopalan, S. S. (2021). Towards Effective Detection of Botnet Attacks using BoT-IoT Dataset. https://scholarworks.rit.edu/theses/10698/

[25] Tang, T. A. et al. (2016). Deep learning approach for Network Intrusion Detection in Software Defined Networking. In the *International Conference on Wireless Networks and Mobile Communications*, 258-263.

[26] Yin, C. et al. (2017). A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. *IEEE Access*, 5:21954-21961.

[27] Abrar, I. et al. (2020). A Machine Learning Approach for Intrusion DetectionSystem on NSL-KDD Dataset. In the *Proceeding of ICOSEC,* 919–924.

[28] Althubiti, S. A., Jones, E. M. and Roy, K. (2018). LSTM for Anomaly-Based Network Intrusion Detection. In the *28th International Telecommunication Networks and Applications Conference*, 1-3.