International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparison of Various Classification Techniques in Cyber Security Using Iot

Urvashi Sangwan[1], Dr Rajender Singh Chhillar[2]
[1]Ph.D Scholar
[2]Professor
Department of Computer Science and Applications
Maharshi Dayanand University, Rohtak
usangwan@gmail.com, chhillar02@gmail.com

Submitted: 22/07/2022 **Accepted**: 25/09/2022

Abstract: The Internet of Things (IoT) devices connected to internet increases rapidly in past decade and expected to add more in coming years. These are small devices and many of them hold personal information saved in it. That’s why it needs to be cyber secure so that attackers or intruder don’t misuse anyone’s useful information. As IoT devices are small in size and the security standards lack here that are applicable for non IoT devices. There is a need to prevent attackers to intrude. In this regard this research paper is an attempt to study machine learning (ML) algorithms that are recently used in securing these devices. In this paper various machine learning classification techniques are used and compared.

Keywords: *Internet of things, Security, machine learning, classification.*

1. Introduction

The Internet of Things (IoT) is small devices connected with internet having sensors. These devices are used to automate certain tasks such as in retail sector, healthcare, Agriculture, Smart cities, smart homes and manufacturing industries. It is estimated that in coming years it has been growing to billions in amount. 79% of small and medium organizations say that they have security issues making them powerless against programmers and the developing cybercrime industry. The financial motivated attacks govern to category of Organized crime boasts 80% the malicious actors. Such a massive device needs to be secure as many devices store information that can be used for malicious activity. Because of limitation of its small size, it is very hard to implement strong security system. For this purpose machine learning is used. Machine learning works as a protective shield for these smart devices. This paper divided into many section. Section I is about machine learning, section II describe cyber attack and section III list out systematic analysis of accuracy in classification techniques for cyber security using IoT by various authors.

1.1 Machine Learning

IoT cyber security needs some technology to keep track of all devices connected to internet. ML can protect IoT devices by automating the scanning and management of IoT devices. They can scan and shut down attack automatically. i.e. shut down a Trojan malware attack [1]. ML not only shut down malware but also detect devices later add on intermittently. It can automate the roll out of network segment by adding devices automatically to appropriate segment based on certain set on rules.

There is various classification techniques used in ML. Supervised learning can be further categorized into classification and regression algorithm. Classification model identifies which category an object belongs to whereas Regression model predicts a continuous output. Sometimes there is an ambiguous line between classification and regression algorithms. Many algorithms can be used for both, and classification is just regression model with a certain threshold. When the number is higher than the threshold it is classified as true while lower than is classified as false.

1.2 Linear and Logistic Regression

In this type of learning machine learns by guidance. Labeled data is given to machine and output is generated by seeing this pattern. Training data is available. Linear regression is utilized in strategies to anticipate subordinate variable (y) in view of upsides of autonomous variable (x). When any problem is continuous then it is best used. It is best use in predicting only two variables which is linearly calculated with other variable. The mathematical formula for linear regression is

$$y = b0 + b1x + e \tag{1}$$

Logistic regression is basic algorithm of classification. It is a type of solving and assuming a dependent variable, given a set of independent variable such as dependent variable is categorical. Categorical variable means either the value is 0/1 and yes/no. it is used to predict binary response variable. The S curve matches the relationship between variables. A threshold value is set to determine the probability of event. The mathematical formula for logistic regression is

$$Log[y/1-y] = c + b1x1 + b2x2 + \cdots. \qquad (2)$$

In python a model can be prepared by importing sklearn library.

Decision Tree: Decision tree fabricates a tree having hubs that branches in a progressive system approach and each branch can be parceled as an if-else proclamation. The branches create by dividing the dataset into subsets in view of most significant highlights. Last characterization occurs at the leaves of the decision tree.

Random Forest: Random Forest is an assortment of numerous choice trees. It is a typical sort of group strategies which total outcomes from various indicators. Random Forest furthermore uses packing procedure that permits each tree prepared on an irregular examining of unique dataset and takes the greater part vote from trees. Contrasted with decision tree, it has better speculation yet less interpretable, in light of additional layers added to the model.

Support Vector Machine (SVM): Support vector machine tracks down the most effective way to characterize the information in light of the situation corresponding to a boundary between certain class and negative class. This line is known as the hyper plane which expands the distance between data of interest from various classes. Like decision tree and Random Forest, support vector machine can be utilized in both classification and regression, SVC (support vector classifier) is for classification issue.

K-Nearest Neighbour (KNN): k nearest neighbor calculation as addressing every data of interest in a n layered space which is characterized by n highlights. Also, it computes the distance between guides one toward another, then appoint the name of unseen information in light of the marks of closest noticed data of interest. KNN can likewise be utilized for building suggestion framework.

Naive Bayes: Naive Bayes is based on an approach to calculate conditional probability in view of earlier information, and the guileless supposition that each component is independent to each other. The biggest advantage of Naive Bayes is that, while most machine learning algorithms rely on large amount of training data, it performs relatively well even when the training data size is small. Gaussian Naive Bayes is another name of Naive Bayes classifier that follows the normal distribution.

1.2 IOT

The concept of IoT (Internet of Things) is no longer a mystery. Recently, it has evolved into a tool that can and will affect the way we live in the future. We humans are naturally inquisitive, and we want to simplify and streamline our lives via more connectivity to one another and less manual labour and the possibility of mistake by using Internet-connected devices (IoT). So, we programmed intelligence into our gadgets and attended to other matters that would boost our efficiency. By linking devices to one another and the internet, we've made it possible for them to collect and share data using Machine Learning and Neural Networks. The outcomes of this phase were outstanding. Users of the IoT may take advantage of enhanced levels of automation, analysis, and integration. They enhance the granularity and breadth of coverage in these regions. There are sensors, networks, and robots that make up the IoT. The IoT makes advantage of cutting-edge software, cheaper hardware, and modern perspectives on the role technology plays in our lives. The use of artificial intelligence, sensors, active involvement, and tiny devices is crucial to the success of the Internet of Things. In fig1, we may get a quick idea of what can be accomplished by implementing each of these traits.

1.3.1 Key Features of IOT

The use of artificial intelligence, sensors, active involvement, and tiny devices is crucial to the success of the Internet of Things. In fig1, we may get a quick idea of what can be accomplished by implementing each of these traits.

1) In context of IoT, data is collected & analyzed using artificial intelligence algorithms and networks to turn seemingly inanimate objects into "smart" ones. A easy option may be to install sensors in your fridge and cabinets that tell you when you're running low on milk or cereal.
2) IoT connection is no longer dependant on only one or two service providers, as a result of new enabling technologies, which is a huge step forward for the networking industry. To connect its many nodes, the IoT essentially builds its own mini-networks.

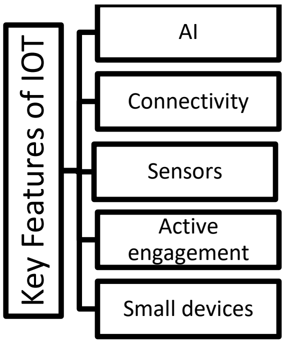

Fig1 Key Features of IoT

3) Without sensors, the Internet of Things loses its special status. With the help of these defining tools, IoT evolves into a functional, adaptable system.
4) It's becoming more unusual for people to interact with technology in a passive fashion. The IoT provides a novel approach of interacting with data, products, and services.
5) There has been a natural progression toward smaller, cheaper, and more powerful electronics. The accuracy, scalability, and flexibility of the Internet of Things (IoT) rely on tiny, purpose-built devices.

1.4 **Cyber Attack in IoT**

1. **Intrusion detection system:** IDS are an network security innovation initially worked for distinguishing weakness that takes advantage of against an target application or PC. Different types of Intrusion Detection Systems (IDS), Signature-based Intrusion Detection Method, Anomaly-based

Intrusion Detection Method, and Hybrid Detection Method.

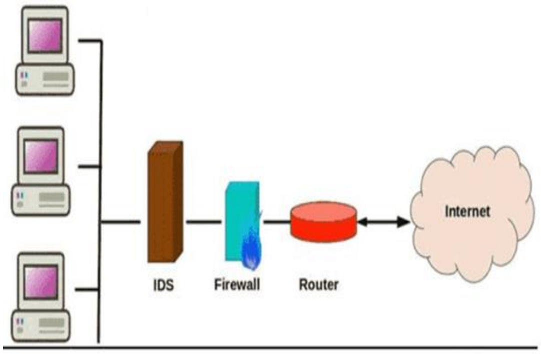

Fig. 2. IDS Architecture

2. **IoT:** Internet of things (IoT) devices is connected to network and each device has identifier that collects and stores information automatically. IoT devices are divided into many parts as consumer application that controls your home AC to close the door. Business application has sensor to check machinery, inventory details and government application use traffic lights, automatic car drive to weather and land control. All these devices need to be secure for intruder to gain useful information and harm the whole system.

1.5 Cyber Security

Cyber security refers to the technologies and process which are designed to protect computer, networks and information from an unauthorized attack. It encloses the risk like hacking, viruses to computer hardware, software and data. Cyber security standards are the standards which allow the organization to practice safe security technique to minimize the number of successful cyber security attacks. There is a vast expansion in the field of computer security is come in to notice. It is because in the present world a large amount of apparatus become internet-enabled and more services move online.

A theft which takes place in the year 2014 in which the top secret information related to Sony Picture's is made public is considered as example of cyber security attacks. By deceiving the safety system of network with the use of malware secret emails, workers data and even unreleased films are hacked.

It is expected for all the association whose work are relies on web that they utilize different classifications of digital wellbeing and safety efforts. It is expected to put a stop on information robbery or disturbance of business. Thusly, as of now or in future the matter of digital protection expert is overwhelmingly popular.

Security-It is the quality or condition secure- - to be liberated from risk. Later on we need to worry pretty much all the security which is examined under security.

Actual Security- In this every issue in which we want to safeguard the actual things, items, and regions of an association from unlawful access and abuse.

Individual security - In this every one of the issues in which we want to safeguard individual or gathering of people who are authorized to get to the association and its activities.

2 Literature Review

Ebu Yusuf Güven et al first normalize data by oversampling an unbalanced no of labels, reducing the size of data set, They used random forest, decision tree, logistic regression and naïve bayes classifier on CICIDS 2017 data set, these classification techniques were compared and in all these random forest had highest accuracy [1]. Mujaheed Abdullahi et. Solve security and privacy problem of IoT device by machine learning and block chain technique. They reviewed previous twelve years of paper from 2008 to 2019 on cyber security using IoT [2]. MILOUD BAGAA et al provide full detail on Denial of service attack to network intrusion and data leakage. This paper introduce security framework for both software defined networking and network function virtualization. This framework enables for combating different threats anomaly detection in IoT. To provide better security combine supervised learning, distributed data mining and neural network. Later conduct experiment on smart building using anomaly based intrusion detection for IoT [3]. Fatima Hussain et al. Reviewed various supervised and unsupervised machine learning models such as SVM, KNN, ANN, naïve bayes, ensemble learning, PCA and k-means clustering. Describe their advantages and disadvantages based on type of attack and need of security. Various applications of machine learning are used in providing security to IoT [4]. Maryam Anwer et al. Provide security to software defined networks and fog layer of network against malicious and anomalous data within IoT systems. They examine various classification techniques on KDD cup 99 dataset. The random forest and k-nearest neigbour are more efficient [5]. Khalid Albulayhi et al. Intrusion detection system using entropy based feature selection. The information gain and gain ratio had selected to come to relevant feature. These selected features undergoes through set theory of union and intersection. The data set used was IoTID 20 and NSL-KDD. Various classification methods are compared to build new model such as bagging, multilayer perception, J48 and IBK [6]. Amine Khatib et al. Anomaly based IDS. First imbalance data transform to balance by SMOTE technique. UNSW-ND-15 dataset used to check performance of various classification techniques. The nystrom based kernel SVM show high accuracy. An evaluation metrics drawn to show accuracy, precision, f1 score and auc-roc curve [7]. Sundar Krishnan et Need to secure industrial IoT, various supervised and unsupervised learning model had used. SVM and random forest show high accuracy. XGBoost, NN and RNN also show promising output [8]. Maryam Anwer et al. IoT network connected to fog to things connections. Need to secure cloud layer, fog layer and terminal layer. The NSL-KDD dataset is used to design model. SVM, Gradient boosted decision tree and random forest showed high accuracy [9]. NAZAR WAHEED et al. Build standard security framework to secure fog and edge computing along with cloud applications [10]. Yakub Kayode Saheed Detect DoS and spoofing attack. SVC, XGBoost and random forest supervised learning technique used to learn different pattern. Here random forest has highest accuracy [11]. Jadel Alsamiri et al. Review various ML and DL approach to provide various security schemes [12]. P. Roshni Mol et al. BoT-IoT data set was used. CICF to extract flow based protocol 25 features had extracted. Compare random forest with KNN [13].

2.1 Below is list of paper published that use machine learning classification for cyber security by using IoT

S no.	Author/year	Title	Classification technique	Accuracy	Limitation
[1]	Ebu Yusuf Güven et al./2022	Multiple Classification of Cyber Attacks Using Machine Learning	Random Forest	99.94%	More work on data pre-processing. A proper model is not formed.
[2]	Khalid Albulayhi et al./ 2022	IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method	Bagging, Multilayer Perception, J48, and IBk	99.98%	Feature selection is done using machine learning. No model developed.
[3]	Yakub Kayode Saheed./2022	A machine learning-based intrusion detection for detecting internet of things network attacks	PCA XGBoost	99.99%	Supervised classification is used.
[4]	Mujaheed Abdullahi et al./2022	Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review.	Supervised learning classification.	Above 99%	A systematic survey from 2008 to 2019
[5]	Shaftab AHMED et al./2022	IoT Based Smart Systems using Machine Learning (ML) and Artificial Intelligence (AI): Vulnerabilities and Intelligent Solutions	All classification techniques	Between 97% to 99%	Only systematic survey.
[6]	Amine Khatib et al./2021	Machine Learning based Intrusion Detection for Cyber-Security in IoT Networks	SVM using SMOTE tcchnique.	94%	The data set for IoT devices are hard to met. The performance is low.
[7]	Maryam Anwer et al./2021	Attack Detection in IoT using Machine Learning	Random forest	85.34%	Only fog layer is identified for attack.
[9]	Sundar Krishnan et al. /2021	IoT Network Attack Detection using Supervised Machine Learning	SVC XGBoost Random Forest	98.20% 99.31% 99.23%	Only do feature selection. No cybersecurity model.
[10]	Rushit dave et al./2021	An analysis of cyber security driven by machine learning.	Random forest	86%	Only a single classification method achieves this accuracy. The level of accuracy is low.
[11]	P. Roshni Mol et al. /2021	Classification of Network Intrusion Attacks Using Machine Learning and Deep Learning.	Adaboost	99.8%	Fail to identify active threats.
[12]	Fatima Hussain et al. /2020	Machine Learning in IoT Security: Current Solutions and Future Challenges	All classification technique	Above 99%	Model trained is application specific.
[13]	NAZAR WAHEED et al./2020	Security and Privacy in IoT Using Machine Learning and Blockchain: Threats and Countermeasures.	Adaboost	99.91%	Privacy preserving data analysis is not present there.
[14]	MILOUD BAGAA et al./2020	A Machine Learning Security Framework for Iot Systems	One class SVM	99.71%	Adding new IoT device means adding new burden to security. The performance going low by adding new devices.
[15]	Jadel Alsamiri et al./2019	Internet of Things Cyber Attacks Detection using Machine Learning	All classification technique	Average 98% to 99%	Only supervised algorithm is evaluated.

Table 1. Literature Survey

3 Problem Statements

Research on IoT security has been done, however it has focused on classic methods. There are, however, several studies focusing on the classification in IoT cyber security. These studies, on the other hand, did not lead to a practical solution. In addition, the standard security system has a performance problem. In order to increase the security of the Internet of Things (IoT), it is necessary to integrate machine learning to security system.

4 Need of Research

The IoT data has not been protected. Efforts are urgently required to develop an advance protection system that provides a strong and secure solution. This proposed technology combines machine learning approach and the Internet of Things with new machine learning model. Finally, a comparison of data security and performance has been made.

5 Conclusions

When comparing all supervised machine learning algorithms such as logistic and linear regression, naïve bayes, KNN, SVM, decision tree, random forest, adaboost in exploring latest paper on cybersecurity by using IoT dataset it is found that adaboost and random forest show high accuracy. It can be further improved by reducing its complexity.

6. References and Footnotes

6.1 References

[1] Ebu Yusuf Güven, Sueda Gülgün , Ceyda Manav , Behice Bakır , Zeynep Gürkaş Aydın. " Multiple Classification of Cyber Attacks Using Machine Learning." Electrica 2022; 22(2): 313-320.

[2] Nazar Waheed, Xiangjian He, Muhammad Ikram, Muhammad Usman, Saad Sajid Hashmi, Muhammad Usman, "Security and Privacy in IoT Using Machine Learning and Blockchain: Threats and Countermeasures" ACM Comput. Surv., Vol. 53, No. 3, Article 1. Publication date: April 2020.

[3] Yakub Kayode Saheed a, Aremu Idris Abiodun b, Sanjay Misra c,*,Monica Kristiansen Holone c, Ricardo Colomo-Palacios "A machine learning-based intrusion detection for detecting internet of things network attacks" Alexandria Engineering Journal (2022) 61, 9395–9409

[4] Garg, D. K. . (2022). Understanding the Purpose of Object Detection, Models to Detect Objects, Application Use and Benefits. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 01–04. https://doi.org/10.17762/ijfrcsce.v8i2.2066

[5] Sam Strecker1, Willem Van Haaften2, and Rushit Dave3 " An Analysis of IoT Cyber Security Driven by Machine Learning" https://www.researchgate.net/publication/354068962

[6] Amine Khatib1,2,3, Mohamed Hamlich1, Denis Hamad2 "Machine Learning based Intrusion Detection for Cyber-Security in IoT Networks " E3S Web of Conferences 297, 01057 (2021)

[7] Mujaheed Abdullahi, Yahia Baashar, Hitham Alhussian, Ayed Alwadain, Norshakirah Aziz, Luiz Fernando Capretz and Said Jadid Abdulkadir, "Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review" Electronics 2022, 11, 198.

[8] Shaftab AHMED, Mohammad ILYAS, M. Yasin Akhtar RAJA "IoT Based Smart Systems using Machine Learning (ML) and Artificial Intelligence (AI): Vulnerabilities and Intelligent Solutions" Proceedings of the 13th International Conference on Society and Information Technologies (ICSIT 2022)

[9] Sundar Krishnan, Ashar Neyaz & Qingzhong Liu, " IoT Network Attack Detection using Supervised Machine Learning", International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (10) : Issue (2) : 2021

[10] Fatima Hussain, Rasheed Hussain, Syed Ali Hassan, and Ekram Hossain," Machine Learning in IoT Security:Current Solutions and Future Challenges". IEEE Communications Surveys & Tutorials · April 2020

[11] Jadel Alsamiri1, Khalid Alsubhi2," Internet of Things Cyber Attacks Detection using Machine Learning", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 10, No. 12, 2019

[12] Kabisha, M. S., Rahim, K. A., Khaliluzzaman, M., & Khan, S. I. (2022). Face and Hand Gesture Recognition Based Person Identification System using Convolutional Neural Network. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 105–115. https://doi.org/10.18201/ijisae.2022.273

[13] P. Roshni Mol, Dr. C. Immaculate Mary," Classification of Network Intrusion Attacks Using Machine Learning and Deep Learning", Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 2, 2021, Pages. 1927 – 1943

[14] Maryam Anwer, Muhammad Umer Farooq, Shariq Mahmood Khan, Waseemullah, "Attack Detection in IoT using Machine Learning", Engineering, Technology & Applied Science Research Vol. 11, No. 3, 2021, 7273-7278

[15] Khalid Albulayhi, Qasem Abu Al-Haija, Suliman A. Alsuhibany, Ananth A. Jillepalli, Mohammad Ashrafuzzaman and Frederick T. Sheldon, "IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method", Appl. Sci. 2022, 12, 5015.

[16] A. Thakkar, and R. Lohiya, "A review of the advancement in intrusion detection datasets," Procedia Comput. Sci., vol. 167, pp. 636–645, 2020.

[17] Malla, S., M. J. . Meena, O. . Reddy. R, V. . Mahalakshmi, and A. . Balobaid. "A Study on Fish Classification Techniques Using Convolutional Neural Networks on Highly Challenged Underwater Images". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 01-09, doi:10.17762/ijritcc.v10i4.5524.

[18] N. Ye, X. Li, Q. Chen, S. M. Emran, and M. Xu, "Probabilistic techniques for intrusion detection based on computer audit data," IEEE Trans. Syst. ManCybern. A Syst. Hum., vol. 31, no. 4, pp. 266–274, 2001

[19] S. Rastegari, P. Hingston, and C. Lam, "Evolving statistical rule sets for network intrusion detection," Appl. Soft Comput., vol. 33, no. C, pp.348–359, 2015.

[20] S. Rajagopal, P. P. Kundapur, and H. K. S., "Towards effective network intrusion detection: From concept to creation on Azure cloud," IEEE Access, vol. 9, pp. 19723–19742, 2021.

[21] Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization,"In Proceedings of the 4th International Conference on Information Systems Security and Privacy, Vol. 1. Funchal, Madeira, Portugal: ICISSP, 2018, pp.108–116

[22] A. Gharib, I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "An evaluation framework for intrusion detection dataset," In International Conference on Information Science and Security (ICISS), Vol. 2016, 2016. Pattaya, Thailand: IEEE Publications, 2016, pp. 1–6.

[23] Ahmed Cherif Megri, Sameer Hamoush, Ismail Zayd Megri, Yao Yu. (2021). Advanced Manufacturing Online STEM Education Pipeline for Early-College and High School Students. Journal of Online Engineering Education, 12(2), 01–06. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/47

[24] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, "Performance analysis of machine learning algorithms in intrusion detection system: A review," Procedia Comput. Sci., vol. 171, pp. 1251–1260, 2020.

Acknowledgements

This research was supported by Dr. Rajender Singh Chhillar, Professor, Maharishi Dayanand University, Rohtak. I thank other Ph.D scholars who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper.

Author contributions

Urvashi[1] Sangwan[1]: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation. Field study.
Dr. Rajender Singh[2] Chhillar[2]: proper Guidance to write and review this research paper. Together with Visualization and Writing-Reviewing.

Conflicts of interest

The authors declare no conflicts of interest.