



Machine Learning with Data Balancing Technique for IoT Attack and Anomalies Detection

Original
Article

Muhammad Asad Arshed^{1,2*}, Muhammad Abdul Jabbar¹, Farrukh Liaquat¹, Usman Mohy-ud-Din Chaudhary², Danial Karim², Hina Alam¹, Shahzad Mumtaz²

¹School of Systems & Technology, University of Management & Technology, Lahore, Pakistan

²Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

*Correspondence: Muhammad Asad Arshed, Email: muhammadasadarshed@gmail.com

Citation | Arshed. A. M, Jabbar. A.M, Liaquat. F, Chaudhary. M.U, Karim. D, Alam. H, Mumtaz. S, "Machine Learning with Data Balancing Technique for IoT Attack and Anomalies Detection". International Journal of Innovations in Science and Technology. Vol 4, Issue 2, 2022, pp: 490-499

Received | May 9, 2022; **Revised** | May 25, 2022; **Accepted** | May 27, 2022; **Published** | May 29, 2022.

Nowadays the significant concern in IoT infrastructure is anomaly and attack detection from IoT devices. Due to the advanced technology, the attack issues are increasing gradually. There are many attacks like Data Type Probing, Denial of Service, Malicious Operation, Malicious Control, Spying, Scan, and Wrong Setup that cause the failure of the IoT-based system. In this paper, several machine learning model performances have been compared to effectively predict the attack and anomaly. The performance of the models is compared with evaluation matrices (Accuracy) and confusion matrix for the final version of the effective model. Most of the recent studies performed experiments on an unbalanced dataset; that is clear that the model will be biased for such a dataset, so we completed the experiments in two forms, unbalanced and balanced data samples. For the unbalanced dataset, we have achieved the highest accuracy of 98.0% with Generalized Linear Model as well as with Random Forest; Unbalanced dataset means most of the chances are that model is biased, so we have also performed the experiments with Random Under Sampling Technique (Balancing Data) and achieved the highest accuracy of 94.3% with Generalized Linear Model. The confusion matrix in this study also supports the performance of the Generalized Linear Model.

Keywords: IoT Attacks, IoT Anomalies, Random Under Sampling, Machine Learning

Project details.

Nil

CONFLICT OF INTEREST:

The author(s) declare that the publication of this article has no conflict of interest.

Author's Contribution. All the authors contributed equally



TOGETHER WE REACH THE GOAL



INTRODUCTION

The Internet seems to be a basic need for people. People use the Internet according to their needs. They use it not just for entertainment but also for other purposes to fulfill daily needs. Almost 48% of people use the Internet [1].

Due to the popularity of the Internet, another field is being emerged, named the Internet of Things (IoT). Only tablets, computers, and mobiles were connected to the Internet in the past. But currently, due to the IoT, Many devices like air conditioning and television are connected to the Internet. IoT is now being used in healthcare, traffic monitoring, and agriculture. It provides intelligent services without any intervention. Sensors are used for data procurement, and data is populated through the network. Privacy is a significant concern. Privacy is a paramount concern during the collection and transmission of the data [2]. Sensors, devices, and nodes are used to form IoT devices that enable communication without the intervention of humans. Different sensors like temperature, position, pressure, proximity, etc., are used in IoT devices. Sensors help control Open/close doors, AC devices, Lights, etc. [3]. Communication protocols and messaging protocols enable communication between different nodes. Node to node messaging protocols include COAP, XMPP, HTTP, AMQP, and MQTT [4].

IoT device sensors use the wireless medium to propagate data, making it easy to attack. Smart sensors typically operate on low energy and provide less support for attacks. Thus, there is always a risk of attacks on IoT devices, such as on-off attacks. Other types of attacks can inject false data packets, which result in the consumption of node energy and network bandwidth [5]. There are two types of IoT attacks, cyber-attacks, and physical attacks. Cyber-attacks affect wireless networks resulting in stealing, changing, damaging, and changing data. On the other hand, physical strikes cause physical devices to harm [6]. Cyber-attacks can be classified into five types. (i) Based on purpose (ii) Legal classification (iii) Based on the severity of involvement (iv) Based on scope (v) Based on the network type. In contrast, five primary security goals are Availability, Integrity, Authentication, Confidentiality, and Non-repudiation [7].

Every field seems to need IOT services nowadays. With the emergence of technologies, attacks have always been a threat to the system as IoT systems are hot nowadays, so attackers are keen to attack and harm the systems. On the other hand, researchers are always there to help the community to restrain these types of attacks.

Inspired by natural intelligence, making machines think like humans is artificial intelligence. AI is divided into two subparts (i) Strong AI and (ii) Weak AI. Strong AI claims that machines will think precisely like human intelligence. In weak AI, the device is trained with pre-existing data. And the machine performs intelligently [8]. Artificial techniques are being widely used to detect these types of attacks. Machine Learning and Deep Learning Neural Networks are the sub-branches of AI facilitating detection. Machine Learning is a subfield of AI consisting of computational algorithms used to inject intelligence into machines inspired by human intelligence. Machine Learning has gained wide acceptance in various fields, including Computer Vision, Pattern Recognition, IoT applications, medical applications, Biomedical, and other areas [9].

Due to the number of connections, IoT plays a vital role in secured communication. The introduction of machine learning can provide adequate privacy and security. It is challenging to evaluate which model is effective for which problem in this environment. Based on types of attacks, Binary, as well as Multiclass attacks, exists there. Different algorithms are used to find binary and Multiclass classification Attacks. K-nearest neighbor, Support vector machine, decision tree, Naïve Bayes, Random Forest, Artificial Neural Network, and logistic regression are used in finding both types of classifications in intrusion detection systems [10]. Machine Learning has opened up new avenues for research. It helps to identify suspicious activities and threats to the system [11].

Literature Review

IoT-based Darknet traffic detection Systems are evaluated and investigated using supervised Machine Learning techniques [12]. The world is changing its direction toward intelligent, innovative technologies such as smart homes, smart cities, and smart industries. With the increasing demand for smart technologies, nodes are rapidly increasing. Hence causes anomalies in the system. The abnormalities in the system are being detected to avoid loss of

data and effective working of smart devices using Machine Learning [13]. Denial of Service attack, spoofing attack, jamming, and eavesdropping attack detections and preventions are operating Machine Learning Techniques, including Supervised Learning, Unsupervised Learning, and Reinforcement Learning. It ensures data privacy for the user [14]. An enhanced encryption method using the AI technique is proposed. The data shared matrix achieves the (k,n) threshold strategy [15]. A real-time solution for SDN based on intrusion detection and mitigation solution using random forest classification that claims high accuracy for attacks in SDN-managed networks [16]. The article proposes a secure control authentication scheme from cloud to end devices, enabling direct secure communication, resource use, and power consumption [17].

. Machine learning identifies the proposed model's accuracy, efficiency, and performance [18]. The formal Modeling technique is used to verify and validate cyber-attacks, and the ML reinforcement approach is used for the cyber-attacks [19]. Artificial immune systems (AIS) are used to overcome security issues. AIS solutions for IoT security issues have been discussed [20]. A deep neural network is used to propose an anomaly detection system in IoT for a secure network. Different deep learning models are applied, such as RNN, CNN, DNN, and some variants [21]. A hybrid algorithm-based new framework model is proposed for the cyber-attack detection system. The proposed model effectively uses a Machine learning algorithm [22]. The article presents an intelligent framework in the combination of Complex event processing technology and machine learning to detect IoT security attacks. The IoT-based healthcare system has been selected to validate the architecture, demonstrating satisfactory results [23].

Table 1: Literature Summary

Author	Model Name	Technique	Result
Liang Xiao et al. [14]	IoT attack models	Supervised Learning, Unsupervised Learning, Reinforcement Learning	Spoofing detection accuracy is 95%, user identification accuracy is 92.34%
Dhanke Jyoti Atul et al. [18]	Energy-Aware Smart Home	J48 tree Classification Algorithm, Naïve Bayes Algorithm	The accuracy rate is 85%
Jose Roldan et al. [23]	An Intelligent Architecture	Linear Regression, Support Vector Machine	0.99998 precision
Zeeshan Ahmad et al. [21]	Efficient Anomaly Detection	1-D Convolutional Neural Network, Recurrent Neural Network, Gated Recurrent Unit, Long Short-term Memory	Model Accuracy 0.57-2.6%, FAR reduces by 0.23-7.98%, detection accuracy 0.99-3.45%
Indrajit Mukherjee et al. [13]	Anomalies Prediction Model	Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Artificial Neural Network	Case1: Logistic Regression Accuracy 99.4%, Case 2: Logistic Regression Accuracy 99.99%,

Obective of the Study:

This study aim at investigation of data stability and concludes that imbalanced datasets can cause overfitting of machine learning models. In such cases models always predict the majority class (even in real world environment for unseen data) . Therefore machine learning models can be modified so that memorization model can be generalized and can be more effective in real world environment. We have considered number of well known machine learning models for experiments, and have performed experiment for un-balanced and balanced dataset. A number of techniques available for data balancing from which we have selected random techniques.

Methodology

The Overall framework of this study can be seen in Figure 1. The initial step of this study is Dataset selection and analysis of the data. The next step is cleaning the data, like dealing with noise-free data. In previous studies, there is no method to deal with Imbalanced

datasets; in this scenario, the proposed model led to an overfitting problem and consistently predicted the majority class, the “normal” class in the considered dataset. Figure 1 is for the summary of the proposed framework in which we have represented all the significant steps of the proposed work. The brief flow of the proposed framework is mentioned in below bullet points.

1. **Dataset:** We have considered the open-source dataset of 8 classes.
2. **Preprocessing:** The dataset consists of null values. Most of the features are not necessary, so we applied to preprocess steps like removing null values records or using average(values).
3. **Data Splitting:** After the clean dataset and feature selection, we divided our dataset into train test folds. Mainly, we divide 80% of the data for training and 20% for testing/ evaluation.
4. **Model Training:** With the train data, we trained our considered models with suitable parameters.
5. **Model Evaluation:** After training, we evaluated our model with test data and evaluation matrices.

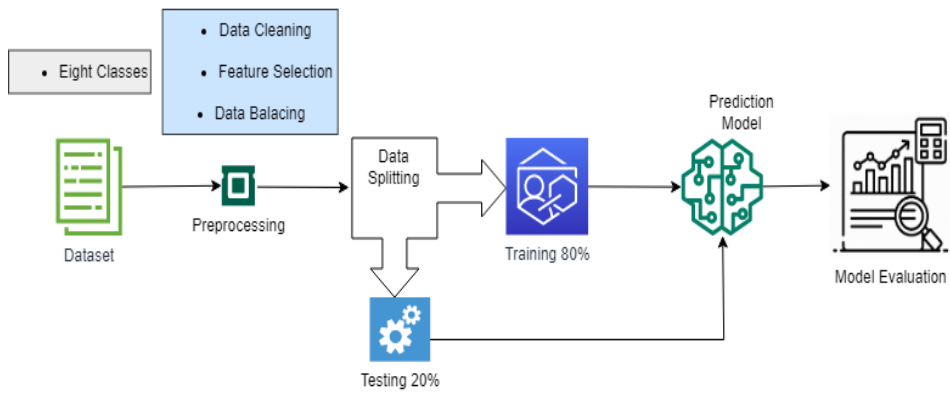


Figure 1: Proposed Framework

Dataset

The dataset is open source and was taken from Kaggle(<https://www.kaggle.com/datasets/francoisxa/ds2ostraffictances>). An IoT-based virtual environment was created to record data using distributed innovative space orchestration system. The communication between the collection of micro-services takes place using a protocol named Message Queuing Telemetry Transport. There were 13 features and eight classes in this dataset. The dataset has missing data 148 and 2050 in terms of “Accessed Node Type” and “Value” features.

The features of the dataset are discussed in Table 2.

Table 2: Features Description

Sr. No.	Feature	Data Type
1	Source ID	Nominal
2	Source Type	Nominal
3	Source Location	Nominal
4	Source Address	Nominal
5	Destination Service Type	Nominal
6	Destination Location	Nominal
7	Destination Service Address	Nominal
8	Accessed Node Type	Nominal
9	Accessed Node Address	Nominal
10	Value	Continuous
11	Timestamp	Discrete
12	Operation	Nominal
13	Normality	Nominal

Data Preprocessing

Some preprocessing is used to make the dataset right for the applied model. In the first step, missing data was handled. "Accessed Node Type" categorizes values while "Value" has continuous values. These two columns contain missing values which results in anomalies in data transferring. "Accessed Node Type" has 148 rows representing values that are 'NaN' (Not a Number). As it is categorical, removing 148 rows might lose valuable data. So, 'NaN' values are replaced by 'Malicious' values in this feature. The "Value" feature also contains some unforeseen data which does not have continuous values. These values are converted into meaningful values. It helps the model to predict better accuracy. The random values "None," "True," "False," and "Twenty" in the "Value" feature are replaced by meaningful values "0," "1," "2," and "20," respectively. Finding feature types in the dataset is a necessary step. The dataset is comprised of Numerical and Categorical data. The depicted feature datatypes are shown in Table 1. can be claimed all features as Nominal except for "Value" and "Timestamp." The "Value" column has continuous values. The "Timestamp" feature is removed as it is not being considered.

The next step is to convert nominal data into vectors. There are many ways to convert categorical data into vectors. Label encoding and one hot encoding schemes are standards. The data is transformed into a feature vector using the label encoding technique. One hot encoding scheme selection might cause an increase in features. Therefore, label encoding is preferred for this dataset.

Feature Selection

We have 12 explanatory variables (features) and 1 response variable in the considered dataset. Feature selection selects a valuable subset of features to develop an effective prediction model. . It improves the model's performance in terms of evaluation scores and is essential to reduce features. Still, it also improves the model's performance in terms of evaluation scores and computational model time and cost decrease.

Currently, statistical models are being used due to effective performance and speed to identify the attributes that have strong relationships with the response/target variable. It is a headache to select which features are reasonable in relation. We can use a built-in tool like Weka and Rapid Minor to solve this problem. In this study, we have considered Rapid Minor to evaluate the relation of features and conclude that all features are essential.

Evaluation Metrics

The current problem is based on a multi-classification task as we have eight classes. For evaluation purposes, we have considered the following metrics.

I. Accuracy

Accuracy measures the percentage of how many predictions are accurate. In other words, if you have a total of 10 samples for testing and seven samples' prediction are correct, then it means the accuracy is 70%.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) * 100 \quad (a)$$

II. Precision

The ratio of accurate prediction and total positive class samples is called precision.

$$Precision = (TP) / (TP + FP) * 100 \quad (b)$$

III. Recall

The ability of the classifiers for accurate prediction out of expected actual.

$$Recall = (TP) / (TP + FN) * 100 \quad (c)$$

IV. F1-Score

The F1-Score can be described as the mean of precision and recall score.

$$F1 - SCORE = 2 * (PR) / (P + R) * 100 \quad (d)$$

Results & Discussion

We have not considered only 1 or 2 models; we have considered the number of well-known machine learning models in this study that are being considered nowadays for different classification and regression problems.

Our problem is a classification problem, and the dataset is imbalanced. The distribution of Samples in terms of classes is shown in Figure 2.

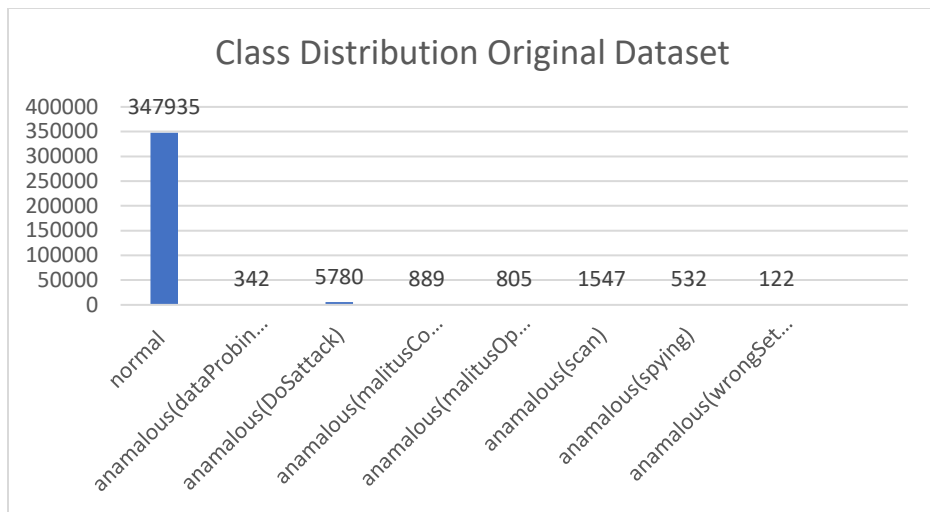


Figure 2: Data Distribution in Terms of Classes in Actual Unbalanced Dataset

Although the dataset is not balanced, we have evaluated the model performance with the original dataset and selected different samples of classes of almost equal data distribution.

Results Comparison and Discussion

Figure 2 determines that it is observable that the dataset is imbalanced, and several studies exist in which results are presented with the imbalanced dataset, which is not suitable for concluding the effective machine learning model. Let's suppose we have 10,000 samples of class A and 200 samples of class B. So, it is common practice that models learn majority class and always predict class A in real life; this problem leads to overfitting. So, to avoid such issues and conclude an effective model with effective predictive ability in real life for any unseen sample, we have presented the result in balanced and unbalanced datasets.

Unbalanced Data Set Results

In this study, we have considered seven well-known machine learning models to compare and find out which model is more effective than others for such a problem. Table 3 has machine learning models' evaluation scores (accuracy) for an unbalanced dataset. The problem here is that when data is unbalanced, model biased or overfitting occurs due to too much learning of majority classes. Although most of the research was published without focusing on this problem, we have presented the result in both forms, balanced and unbalanced dataset results.

Table 3: Evaluation Scores of 7-Machine Learning Models for Unbalanced Dataset

ML Models	Accuracy
Generalized Linear Model	98.0%
Random Forest	98.0%
Logistic Regression	96.1%
Fast Large Margin	97.2%
Decision Tree	71.8%
Naive Bayes	97.2%
Support Vector Machine	97.2%

Balanced Data Set Results (Random Under Sampling)

Under-sampling is a technique in which we balance the dataset to reduce the majority class; for example, we have 2-classes (Male & Female). The number of samples of the Male class is 2000, and the Female class samples are 500. So, to balance the classes, there are two ways, one is to increase the samples of the Female class (Minority Class); that process is called oversampling, and the second way is to reduce the samples of the Male class (Majority Class), and this process is called Under Sampling. To mitigate the problem of model biased / overfitting, we need to consider at least reasonable data samples of each class. We have considered max 300 samples from each class. If any class has samples up to 300, then all the samples of that class are considered, as shown in Figure 3.

The dataset is now in a balanced format in the undersampling technique. Let's now evaluate the model performance (accuracy) for these balanced samples see Figure 4.

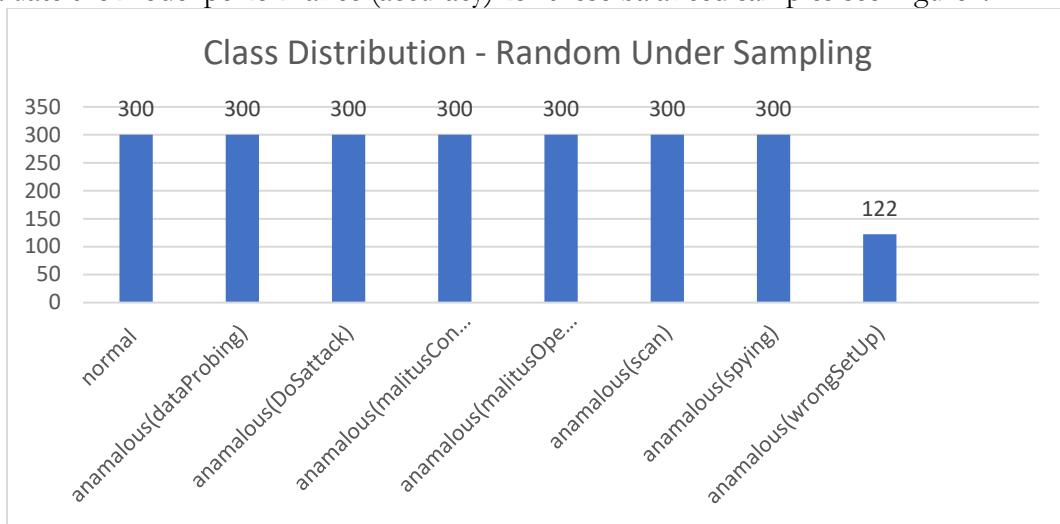


Figure 3: Data Samples Distribution in terms of Classes (Random Under Sampling)

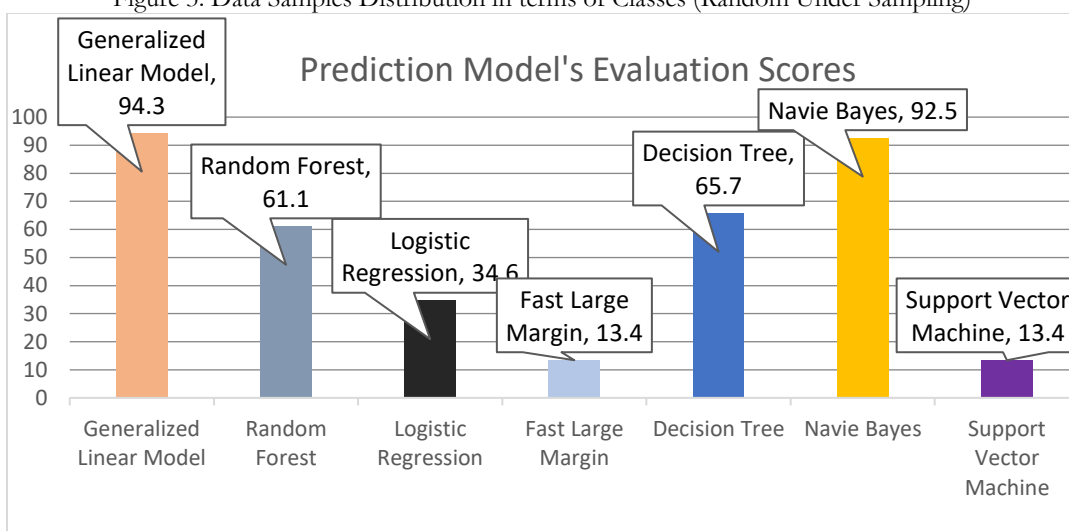


Figure 4: Model Accuracy Comparison of Balanced Dataset (Random Under Sampling)

The machine learning model's performance (Generalized Linear Model) is more effective than other considered models for balanced datasets (Random Under Sampling). To support this statement, we have drawn a confusion matrix seen in Table 4.

Table 4: Confusion Matrix of Generalized Linear Model for Balanced Dataset

	Normal	Anomalous (scan)	Anomalous (malitiousOperation)	Anomalous (DoSattack)	Anomalous (spying)	Anomalous (dataProbing)	Anomalous (wrongSetUp)	Anomalous (malitiousControl)	Class Precision
Normal	86	2	6	5	1	19	0	0	72.27%
Anomalous (scan)	0	84	0	0	0	0	0	0	100.00%

98.77%	100.00%	100.00%	97.01%	100.00%	100.00%	100.00%
0	0	0	0	0	86	100.00%
0	0	0	0	35	0	100.00%
1	0	0	65	0	0	76.47%
0	0	85	0	0	0	98.84%
0	78	0	2	0	0	91.76%
80	0	0	0	0	0	93.02%
0	0	0	0	0	0	97.67%
0	0	0	0	0	0	100.00%
Anomalous (maliciousOperation)	Anomalous (DoSattack)	Anomalous (spying)	Anomalous (dataProbing)	Anomalous (wrongSetUp)	Anomalous (maliciousControl)	Class Recall

From Table 4, it can easily conclude that the Generalized Linear model performance is effective for all the considered classes.

Conclusion

Based on this study, we can say that the data balancing technique is more critical in an analysis where data is too imbalanced, and accuracy is affected by these unbalancing. Whenever the data is unblanched, the model gives you the highest accuracy score due to overfitting, which means the model is trained only for the majority class, and no proper training takes place for all classes. We have performed the experiments for balanced and unbalanced dataset classes in this study to support the concept of how it is essential to consider a balanced dataset. The Generalized Linear Model is effective for unbalanced and balanced (Random Under Sampling) classes for such a task, but the evaluation scores are different for the balanced and unbalanced dataset. In this study, we have applied Random Under Sampling and now have planned to use Oversampling technique in the future to balance the dataset for such task.

References

- [1] M. Burhan, R. A. Rehman, B. Khan, and B. S. Kim, "IoT elements, layered architectures and security issues: A comprehensive survey," *Sensors (Switzerland)*, vol. 18, no. 9, pp. 1–37, 2018, doi: 10.3390/s18092796.
- [2] M. A. Rahman and A. T. Asyhari, "The emergence of internet of things (IoT): Connecting anything, anywhere," *Computers*, vol. 8, no. 2, pp. 8–11, 2019, doi: 10.3390/computers8020040.
- [3] D. Sehrawat and N. S. Gill, "Smart sensors: Analysis of different types of IoT sensors," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, no. Icoei, pp. 523–528, 2019, doi: 10.1109/ICOEI.2019.8862778.
- [4] T. M. Tukade and R. M. Banakar, "Data Transfer Protocols in IoT-An Overview," *Int. J. Pure Appl. Math.*, vol. 118, no. 16, pp. 121–138, 2018.
- [5] X. Liu, Y. Liu, A. Liu, and L. T. Yang, "Defending ON-OFF attacks using light probing messages in smart sensors for industrial communication systems," *IEEE Trans. Ind. Informatics*, vol. 14, no. 9, pp. 3801–3811, 2018, doi: 10.1109/TII.2018.2836150.
- [6] S. M. Tahsien, H. Karimipour, and P. Spachos, "Machine learning based solutions for security of Internet of Things (IoT): A survey," *J. Netw. Comput. Appl.*, vol. 161, no. March, 2020, doi: 10.1016/j.jnca.2020.102630.
- [7] M. Uma and G. Padmavathi, "A survey on various cyber attacks and their classification," *Int. J. Netw. Secur.*, vol. 15, no. 5, pp. 390–396, 2013.
- [8] A. S.-S. [ETEBMS-2016] and undefined 2016, "Applications of artificial intelligence & associated technologies," test.globalinfocloud.com, Accessed: May 18, 2022.
- [9] I. El Naqa and M. J. Murphy, "Machine Learning in Radiation Oncology," *Mach. Learn. Radiat. Oncol.*, pp. 3–11, 2015, doi: 10.1007/978-3-319-18305-3.
- [10] M. Abdullahi *et al.*, "Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review," *Electron.*, vol. 11, no. 2, pp. 1–27, 2022, doi: 10.3390/electronics11020198.
- [11] S. H. Haji and S. Y. Ameen, "Attack and Anomaly Detection in IoT Networks using Machine Learning Techniques: A Review," *Asian J. Res. Comput. Sci.*, no. June, pp. 30–46, 2021, doi: 10.9734/ajrcos/2021/v9i230218.
- [12] Q. A. Al-Haija, M. Krichen, and W. A. Elhaija, "Machine-Learning-Based Darknet Traffic Detection System for IoT Applications," *Electron.*, vol. 11, no. 4, 2022, doi: 10.3390/electronics11040556.
- [13] I. Mukherjee, N. K. Sahu, and S. K. Sahana, "Simulation and Modeling for Anomaly Detection in IoT Network Using Machine Learning," *Int. J. Wirel. Inf. Networks*, no. 0123456789, 2022, doi: 10.1007/s10776-021-00542-7.
- [14] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security?," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 41–49, 2018, doi: 10.1109/MSP.2018.2825478.
- [15] B. Li, Y. Feng, Z. Xiong, W. Yang, and G. Liu, "Research on AI security enhanced encryption algorithm of autonomous IoT systems," *Inf. Sci. (Njy)*, vol. 575, pp. 379–398, 2021, doi: 10.1016/j.ins.2021.06.016.
- [16] A. K. Sarica and P. Angin, "Explainable security in SDN-based IoT networks," *Sensors (Switzerland)*, vol. 20, no. 24, pp. 1–30, 2020, doi: 10.3390/s20247326.
- [17] T. K. Dang, C. D. M. Pham, and T. L. P. Nguyen, "A pragmatic elliptic curve cryptography-based extension for energy-efficient device-to-device communications in smart cities," *Sustain. Cities Soc.*, vol. 56, p. 102097, 2020, doi: 10.1016/j.scs.2020.102097.

- [18] D. Jyoti, R. Kamalraj, G. Ramesh, K. S. Sankaran, S. Sharma, and S. Khasim, "Microprocessors and Microsystems A machine learning based IoT for providing an intrusion detection system for security," *Microprocess. Microsyst.*, vol. 82, no. November 2020, p. 103741, 2021, doi: 10.1016/j.micpro.2020.103741.
- [19] J. A. Bland, M. D. Petty, T. S. Whitaker, K. P. Maxwell, and W. A. Cantrell, "Machine Learning Cyberattack and Defense Strategies," *Comput. Secur.*, vol. 92, p. 101738, 2020, doi: 10.1016/j.cose.2020.101738.
- [20] S. Aldhaheri, D. Alghazzawi, L. Cheng, A. Barnawi, and B. A. Alzahrani, "Artificial Immune Systems approaches to secure the internet of things: A systematic review of the literature and recommendations for future research," *J. Netw. Comput. Appl.*, vol. 157, p. 102537, 2020, doi: 10.1016/j.jnca.2020.102537.
- [21] Z. Ahmad *et al.*, "Anomaly detection using deep neural network for iot architecture," *Appl. Sci.*, vol. 11, no. 15, 2021, doi: 10.3390/app11157050.
- [22] M. Shafiq, Z. Tian, Y. Sun, X. Du, and M. Guizani, "Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city," *Futur. Gener. Comput. Syst.*, vol. 107, pp. 433–442, 2020, doi: 10.1016/j.future.2020.02.017.
- [23] J. Roldán, J. Boubeta-Puig, J. Luis Martínez, and G. Ortiz, "Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks," *Expert Syst. Appl.*, vol. 149, 2020, doi: 10.1016/j.eswa.2020.113251.



Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.