JIRBDAI

Journal of Institutional Research, Big Data Analytics and Innovation

Towards Building an Improved IoMT Attack Classification Model-A Methodological Framework and Analysis of the Experimental Dataset

Saka B. A., Oyelakin A. M.

Department of Computer Science, Crescent University, Abeokuta, Nigeria
Correspondence Email: moruff.oyelakin@cuab.edu.ng

ABSTRACT

Internet of Medical Things (IoMT) is currently revolutionizing the healthcare industry as it can help in the advancement of peoples' health condition as well as productivity of health professionals in service delivery. Despite the positive developments in the use of technology in health sector, people with malicious intent have been launching attacks on heterogeneous medical devices and networks. Machine learning (ML)-based models have been proven to have the ability to intelligently analyze huge volume of IoMT traffic and then classify attacks. To effectively detect zero-day attacks in IoMT with reduce rate of false positives using of ML-based methods, there is a need to address different issues such as high class imbalance, feature selection, and multi-class classification in experimental datasets. This study proposes a multi-stage methodological framework that can be used to adequately classify multi-class cyber-attacks using a recent CICIoMT2024 security dataset. The framework details the various stages that the proposed ML-based attack classification models will contain by taking some of the issues identified in the exploratory data analysis (EDA) layer into considerations. Lastly, Aquila Optimizer is suggested to be used for the selection of most promising attributes while Tree and Non-Tree based ensemble learning algorithms are proposed for the attack classification task. It is concluded that this methodological framework can be used for building improved multi-class attack classification models in IoMT scenario.

KEYWORDS: Feature Selection, Model Performance, Internet of Things, Smart Healthcare

I. INTRODUCTION

Internet has made the global space to be digitally connected and this has brought about generation of wealth of data in different hemisphere of our lives. The health sector has also benefited from the prevalence of internet and computing devices. Ajagbe et al. (2024) mentioned that Internet of Medical Things (IoMT) has drastically brought about cost reduction in healthcare services as well as improvement in user's experience and patient's well being. Healthcare sector is one of the major domains where wealth of sensitive and non-sensitive data is being generated on daily basis with the help of smart medical devices. IoMT has made medical care to be more personalize and accessible. It has revolutionized the whole processes of medical activities. These processes include taking patients' vital signs, diagnosis, test examination and interpretation of result, administering drugs and effective treatment plan. The adoption of IoMT has significantly improved patients' live care, healthcare service accessibility and remote monitoring (Pelekoudas-Oikonomou et al., 2022). The technology, generally, provides more avenues for researcher in using the huge volume of data generated in the landscape to indicatively derive hidden knowledge and insight to develop innovative ideas, that proffer solutions to many problems such as data privacy and security. IoMT comprises of complex architecture which involves heterogeneity of various bio medical smart devices such as wearable, implantable and autonomous devices using diverse protocols for data collection, transmitting and processing (Pradyumna et al., 2024). However, despite many benefits provided by IoMT, it is constrained by data security and privacy challenges. Attackers are launching attacks on smart healthcare infrastructure because of the huge volume of generated data which can be of economic benefits to them (Bushan et al., 2023).

Despite the use of ML techniques for attack classification in IoMT, there is need for innovative ML-based approach to address building models that cannot generalize well. This study proposes a multi-

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

stage methodological framework that can be used to attain improved multi-class attack classification from the chosen dataset. The first task is to present the methodological framework and then briefly explain the various layers it contains. These layers include extensive exploratory analysis of the dataset, data cleaning, and a mention of meta-heuristic based feature selection called Aquilla Optimiser. The framework also mentioned that tree and non-tree ensemble learners will be used for the attack classification model in future study by the authors. The framework is intended to guide a full implementation of the various modules in a choice Python programming language environment by the researchers in subsequent research.

II. LITERATURE REVIEW

Many authors have tailored their worked in the direction of ensuring security and privacy of data in IoMT landscape (Hady et al., 2020; Gopinath & Sethuraman, 2023). This is due to the fact that, IoMT consists of vast array of data from numerous devices with different protocols that are complex in nature. It can be argued that this particular nature makes IoMT environment become the target of cyberattacks. Gopinath et al. (2025) built an ensemble-based binary classification ML model for attack detection in IoMT. The proposed model was built using WUSTL-EHMS-2020 and TON-IoT datasets for effective training and testing phases. The feature selection approach adopted was principal components analysis (PCA) algorithm for feature optimization. More so, the framework leveraged on the decision boundary precision of SVM and combined it with the advantage of XGboost's gradient-based optimization. The authors stated that the model outperforms the other baseline classifiers with accuracy rate of 94.29%. However, the two datasets were not properly analyzed. Any dataset, either real-life data or synthetic data rarely comes clean, hence, the need for thorough exploratory analysis. This is because the fundamental open problem associated with any dataset need to be resolved before using it to develop ML-based model. Alrefaei and Ilyas (2024) proposed a ML-based multi-class classification technique to detect IoT Attacks in a real-time mode. In their study, PySpark architecture was leveraged on using One Vs Rest (OVR) technique. The study utilized the IoT-23 dataset with some ML algorithms such as decision tree (DT), random forest (RF), logistic regression (LR), XGBoost (XGB) and K-nearest neighbour (KNN). Thorough exploratory data analysis was carried out on the dataset to identify the associated open problems with the dataset. As a result of this, data pre-processing which involved data transformation, scaling and synthetic minority oversampling technique (SMOTE) were applied to prepare the dataset for model development. Feature selection was carried out using three different methods. Moshin and Akinul Islam (2024) made use of CICIoMT 2024 security dataset as a pivotal resource with some ML algorithms such as AdaBoost, Random Forest (RF), K-Nearest Neighbors (KNN) and XGBoost in developing effective machine learning (ML) based models for the three different level classification of the dataset. The authors adopted SMOTE for resolving class imbalance. However, no method was used for feature selection.

Pektas et al. (2017) have emphasized the importance of feature selection in any ML-based problem. Thus, it can be argued that the selection of promising features when using a comprehensive dataset like CICIoMT 2024 to build models is of topmost importance to reduce the training time, model complexity and enhance the reliability of the model (Oyelakin et al., 2023). Moreover, selecting good promising network features will enhance the rate of attack detection in IoMT maximally (Narang et al., 2013). Many researchers in the area of classification problem have established feature selection as a pivotal stage in predictive analytics. Oyelakin and Jimoh (2020) built a methodological framework to improve botnet attack detection in cyber space. The authors made use of CTU-13 dataset to build a methodological framework in order to improve botnet attack detection in cyber space. They argued that CTU-13 dataset is the reference botnet dataset that has been identified to be large and realistic. However, the dataset is high in dimension and largely class imbalanced. In their proposed framework, they thus applied SMOTE to solve problem of imbalance associated with the dataset and a filter-based multivariate analysis was proposed for feature selection.

III. METHODOLOGY

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

A. Datasets in IoMT Attack Classification

There are various datasets that have been used so far in the area of cyberattacks detection in IoMT that are publicly available. Some of them include: WUSTL EHMS 2020 Dataset by Hady et al. (2020), ECU-IoHT dataset by Ahmedet al. (2021), ICU dataset by Hussainet al. (2021), CIC-IoT2023 dataset bu Jony and Arnob (2024), as well as the recent dataset named CICIoMT 2024 by Dadkhahet al. (2024). The review paper carried out by Saka et al. (2025) was able to discuss some of the approaches used by authors while building models from the relevant datasets.

B. The Proposed Methodological Framework

The framework contains the various stages that can be used for achieving better multi-class attack classification models in future study. For the purpose of experimentations in further studies, Python IDEs named Spyder and Jupyter notebook will be used. Figure 1 is used to describe the proposed methodological framework in this study.

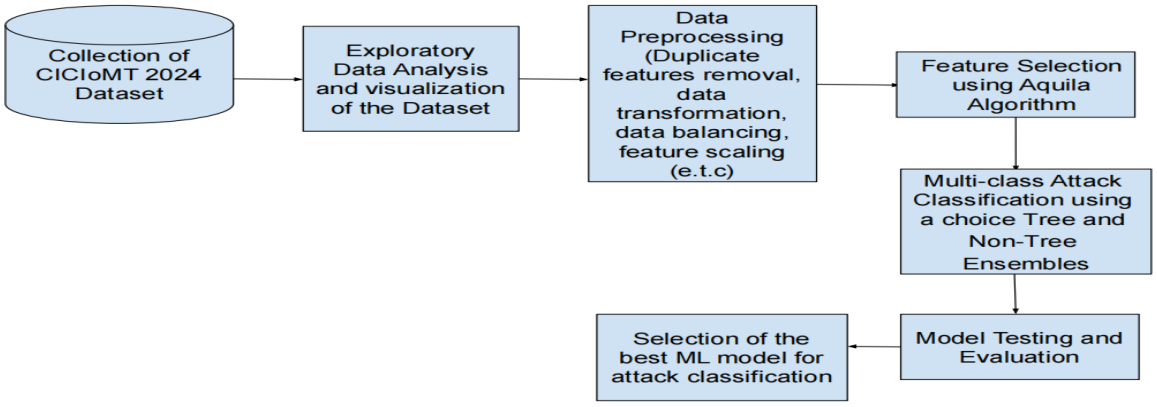

Figure 1: Proposed Methodological Framework

C. Collection of CICIoMT 2024 Dataset

The dataset chosen in this proposed study is CICIoMT 2024 dataset. It was collected from the Canadian Institute of Cyber Security Data Sets repository. It is available for download at https://unb.ca/cic/datasets/iomt-dataset-2024.html as released by Dadkhah et al. (2024). The justifications for chosen this dataset are, its hugeness, recency, relevance and well representation. The dataset is good for evaluating classification of attacks in IoMT landscape with the use of ML-based techniques. The attacks in the dataset are categorized into five. They are: DOS, DDoS, MQTT, spoofing, and reconnaissance with benign. The dataset captures the diversity and intricacy of healthcare's digital infrastructure as argued by Akinul Islam et al. (2024).

D. Detailed Exploratory Analysis

The EDA is part of the module in the methodological framework. The exploration of the chosen dataset will reveal the patterns in the dataset. The exploratory analysis will provide information about the strengths and the problems in the chosen dataset and hence how to better address the open challenges. The various EDA techniques are necessary to reveal the patterns in the dataset in order to make informed decisions about its pre-processing, feature selection and model building and evaluation. The EDA process always helps in identifying presence of duplicate samples, missing values if there is any, outliers, correlations, high class imbalance if present. Adequate understanding of this process will help in building robust ML models with improve performance in line with the argument of Oyelakin et al. (2023) who mentioned that EDA is used to reveal the intrinsic relationship between the input variables and the target class. The most significance of EDA is to avail us with in depth insight into the structure, nature and characteristics of the dataset so as to identify the best ways to use them to achieve robust and reliable ML based attack detection models in IoMT devices. Gibson and Freitas (2015) also stated that EDA is a pivot step in solving any problem related to data analytics as some major open problems will be revealed at this stage. Moreover, the efficiency and effectiveness of any ML-based solution

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

largely depends on the characteristics and nature of data together with the performance of the ML-classifier (Sarker, 2021).

E. Data Cleaning

As contained in the methodological framework, the dataset needs to be pre-processed as identified in figure 1. The decision as to how to clean the data is obtained based on the findings in the EDA earlier carried out.

F. Feature Selection using Aquila Optimization Algorithm

Based on the methodological framework shown in figure 1, the chosen feature selection method to be used at the implementation stage is Aquila Optimizer, a very popular and recent meta-heuristic optimisation algorithm (Abualigah, 2021). The choice is to ensure that the most promising attributes that will help achiev best classification results are selected in the larger CICIoMT2024 dataset. In any classification or regression machine learning model, feature extraction or selection are very important (Kotsiantis et al., 2006; Pektas, 2017). Feature selection generally plays a vital role in building machine learning models. It enhances attack detection by eliminating irrelevant and redundant features, allowing the model to focus on the most significant attributes that distinguish normal and malicious activities. This improves detection accuracy, reduces computational complexity, and speeds up real-time processing, making the model more efficient for resource-constrained IoMT devices (Lispa et al., 2025). Aquila algorithm has what we call convergence analysis which evaluates the ability of an optimisation algorithm to reach a stable and optimal solution within a finite number of iterations. In future study, efforts will be made to ensure that the convergence analysis is adequately taken care of.

G. Proposed Attack Classification Models based on the Selected ML Algorithms

Tree and Non-Tree Classification algorithms will be used for multi attack classification metrics. This is the layer in the methodological framework that will be responsible for the classification of attacks in the dataset. The various algorithms to be used in future study will be investigate as how best they can be used for classifying all the attacks in the data using the identified standard metrics in ML modelling.

IV. RESULTS OF DATASET EXPLORATORY ANALYSIS

The results of Exploratory Data Analysis (EDA) and Visualization carried out in this study are discussed in this section. The EDA is a stage in the proposed methodological framework. Table 1 shows the description of features and the actual sample size of the chosen dataset.

Table 1: Description of the features and sample size in CICIoMT 2024 Training dataset

S/N	Feature Name	Data type
0	Header_Length	float64
1	Protocol Type	float64
2	Duration	float64
3	Rate	float64
4	Srate	float64
5	Drate	int64
	...	...
6	fin_flag_number	float64
37	Tot size	float64
38	IAT	float64
39	Number	float64
40	Magnitue	float64
44	Weight	float64
45	label	object
(7160831, 46)		

dtypes: float64(44), int64(1), object (1)

Table 1 showed that there are forty-four features that are of floating data typs, one inteer data type and one that is categorical (object) in type. During the EDA, it was also discovered that the dataset consists

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

of duplicates rows. Hence, there is a need to remove the duplicates in order to increase the model accuracy when built from the data. The result is demonstrated in the Table 2. It shows the number of duplicated samples in the dataset. Moreover, out of 46 features (columns), it was observed that 45 are numerical datatype (int and float), while one which is the target class is categorical (object) datatype. Thus, the EDA showed us that there is a need to convert the categorical data to numerical datatype for proper processing by the ML algorithms.

Table 2: Description of Duplicate Rows (instances) in Training Dataset

S/N	Header_Length	Protocol Type	...	Weight	label
413640	0.0	1.0	...	141.5	TCP_IP-DDoS-ICMP6_train
413641	0.0	1.0	...	141.5	TCP_IP-DDoS-ICMP6_train
413642	0.0	1.0	...	141.5	TCP_IP-DDoS-ICMP6_train
413643	0.0	1.0	...	141.5	TCP_IP-DDoS-ICMP6_train
413644	0.0	1.0	...	141.5	TCP_IP-DDoS-ICMP6_train
...	...	...	...		...
7160825	54.0	6.0	...	141.5	TCP_IP-DDoS-TCP3_train
7160826	54.0	6.0	...	141.5	TCP_IP-DDoS-TCP3_train
7160827	54.0	6.0	...	141.5	TCP_IP-DDoS-TCP3_train
7160828	54.0	6.0	...	141.5	TCP_IP-DDoS-TCP3_train
7160829	54.0	6.0	...	141.5	TCP_IP-DDoS-TCP3_train

[2645426 rows x 46 columns]
number of duplicated rows: is 2645426

More so, through EDA, it was found that there were no feature mismatches between the training dataset and test dataset. It was also discovered that no missing value or no null values in training dataset. Since, there is no presence of null values, there is no need to handle missing values with any of the popular approaches. Table 3 is used to illustrate that the dataset has no missing or no null values.

Table 3: Result of Checking for missing value in the Training feature Dataset

Feature Name	Number of missing value
Header_Length	0
Protocol Type	0
Duration	0
Rate	0
Srate	0
Drate	0
fin_flag_number	0
...	….
Radius	0
Covariance	0
Variance	0
Weight	0
label	0

dtype: int64

Figure 2 also reflects that The EDA carried out equally showed that the CICIoMT2024 dataset has high class imbalance problem. Figure 2 shows how the dataset is skewed towards majority attack class. Hence, there is need to balance the dataset to capture the minority class to avoid useful information loss or model being biased during experimentation. Synthetic Minority Sampling Technique will be adopted to resolve the issue of class imbalance in future study.

A. Statistical Summary of the Data Set

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

In furtherance to EDA of the dataset, statistical summary of the dataset was critically examined so as to identify the distributions in the training dataset, better. The summary statistics of the captures in the training dataset are shown in Table 4.

Table 4: Results of Summary Statistics of Numerical features of Training Dataset

	Header_Length	Protocol Type	...	Variance	Weight
count	7.160831e+06	7.160831e+06	...	7.160831e+06	7.160831e+06
mean	2.958836e+04	8.042994e+00	...	9.148464e-02	1.414795e+02
std	2.762618e+05	6.292249e+00	...	2.327133e-01	2.174133e+01
min	0.000000e+00	0.000000e+00	...	0.000000e+00	1.000000e+00
25%	5.400000e+01	1.160000e+00	...	0.000000e+00	1.415000e+02
50%	1.080000e+02	6.000000e+00	...	0.000000e+00	1.415000e+02
75%	1.958050e+04	1.700000e+01	...	0.000000e+00	1.415000e+02
max	9.895636e+06	1.700000e+01	...	1.000000e+00	2.446000e+02

[8 rows x 45 columns]

Since the EDA revealed that there are duplicate values, it is therefore required that while using the CICIoMT2024 dataset for attack classification building during further experiments, the redundant or duplicated samples from dataset should be cleaned. This is because duplicate are not adding any new knowledge to the model, rather consuming the computational resources which may impact negatively the training time of the model. Aghabagherloo et al. (2025) as well as Lee et al. (2021) have pointed out that duplicate samples can hinder the performance evaluation metrics of model thereby negatively affect the model generalization. Keeping one representative sample preserves the genuine pattern while getting rid of artificial repetition (Han et al., 2011). Table 5 shows the summary of sample size after removing the duplicates in the dataset.

Table 5: Summary of Number of Rows (instances) before and after Cleaning

Operations	Training Dataset
Number of Rows before cleaning:	7160831
Number of duplicated rows before cleaning:	2645426
Duplicates rows removed with keeping the first one. New shape:	(4515405, 46)
Number of duplicated rows after cleaning:	0
Number of Rows after removing the duplicate:	4515405

B. Duplicate Removal Should Precede Class Imbalance Problem Solving

In addition, it is paramount to understand that duplicate removal operation will precede resolving class imbalance. This is because SMOTE assumes data uniqueness and amplifies any redundancy present in the dataset. Application of SMOTE before duplicate removal leads to synthetic redundancy, overfitting and invalid performance evaluation as argued by Garcia et al. (2019).

C. Attack Categories in the Dataset

Figure 2 is used to provide the visualisation on the attack categories in the dataset. It also includes the distribution of specific attack class in the dataset. The pieces of information gathered herein can further inform the decision to be taken by researchers using the dataset for ML or DL-based model building.

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

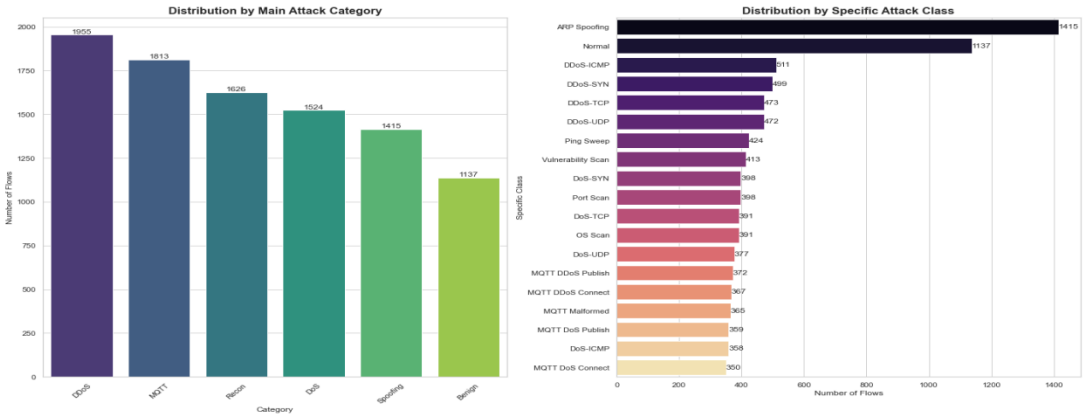

Figure 2: Demonstration of Main Categories and specific types of Attack Class in CICIoMT2024 Training Dataset

D. High Class Imbalance Issues in the Dataset

The EDA in the study also revealed that the chosen dataset has high class imbalance and it will be required to do class balancing prior to using the dataset in experimentation. In the proposed framework, Synthetic Minority Over Sampling Technique that was proposed by Chawla et al (2002) is suggested herein.

DISCUSSION

This study proposed a multi-stage methodological framework that can help in the design of improved multi-class attack classification models in IoMT environment. The framework contains different stages which are considered very important towards building intelligent models that can classify multiple attacks in the candidate CICIoMT 2024 dataset better. The study emphasized the significance of carrying out extensive data exploration which precede the pre-processing stages such as reducing duplicate feature, standardization, feature scaling are proposed, prior to using the model for classification are also recommended. In the study, selection of optimal features using Aquila algorithm that can aid the detection of multi-class attack classification in IoMT is equally emphasized. It is believed that this approach will go a long way in to achieve improved Machine learning-based multi-class attack classification and detection schemes. The EDA carried out revealed that the dataset has duplicate values running into millions in the training dataset and high class imbalance problem. According to Oyelakin et al. (2020), a lot of real-life dataset are associated with open problem of class imbalance and this study further emphasized the need to solve class imbalance issue in ML-based studies. Similarly, Khan and Alkhathami (2024) mentioned that class imbalance can cause bias in the generated solutions. To address extreme class imbalance issue in ML tasks, many research works have adopted different techniques in resolving the problem of imbalance. They include re-sampling techniques such as oversampling of class and under-sampling of majority class (Devi et al., 2021); synthetic data generation techniques like SMOTE (Chawla et al., 2002), Adaptive Synthetic Sampling (ADASYN) (He et al., 2008), cluster-based techniques (Yen & Lee, 2009) and many more. In this propose framework, synthetic minority over-sampling technique (SMOTE) will be considered to resolve the imbalance issue. SMOTE creates artificial examples of minority class by interpolating from nearest neighbour. Khan and Alkhathami (2024) pointed out that the target is to use the process to improve the learning algorithm's performance while reducing the probability of biased model.

In line with the statement of Kotsiantis (2006), the samples of the minority class will be increased through this approach. Mathematically, SMOTE is expressed as in equation 1 as follows:

$$\text{Synthetic\_sample} = x + \lambda \cdot (\text{neighbor} - x) \quad (1)$$

where, x is the original minority class instance. neighbour is one of the k nearest neighbors of x within the minority class. λ is a random value between 0 and 1, controlling the amount of interpolation.

Volume 2, Issue 2, 2026
https://doi.org/10.5281/zenodo.20376497

The result of the EDA also shows that there is a need to do data encoding for the categorical feature in the dataset when it is time for model building. This can be achieved by label encoding depending on the nature of the data. As pointed out by Alrefaei and Ilyas (2024), the encoding involves feature data transformation by altering the structure of data so as to enhance its suitability for analysis and modeling. In the proposed methodological framework, the pre-processed dataset will be fed into Tree-based and Non-Tree based ensembles for the classification of multiple attacks in the dataset.The target is to obtain the optimal results across all the selected metrics in the two different classification scenarios.

V. CONCLUSION

This study introduced IoMT and its roles in the improvement of healthcare delivery. The work further mentioned the need to use intelligent models to identify attacks as attackers are now attacking IoMT infrastructure owning to the prevalence of internet and the high traffic being generated from the smart healthcare. Thus, an argument was made regarding the need for continuous efforts of researchers to keep devising mechanisms to secure smart health care facilities the more. Then, a recent large dataset that is popular for benchmarking ML-based attack identification model in IoMT environment is discussed. The dataset is named CICIoMT2024 and it was critically studied for a better understanding. It is believed that promising attack classification models can be built from the multi-stage intelligent methodological framework. Furthermore, detailed exploratory analysis that showed different data patterns in the dataset was carried out. It was mentioned that the analysis provided actionable insights on what need to be done on the dataset towards ensuring that best attack classification models that will generalize well with reduced false positive rate are achieved in the future work. The study further established that the results of the exploratory analysis will guide other researchers working in similar area, using the dataset. The future work by the authors will focus on using the methodological framework to build and evaluate tree and non-tree machine learning-based models for the effective identification of attacks in the labeled IoMT dataset.

REFERENCES

Abualigah, L., Yousri, D., Abd Elaziz, M., Ewees, A. A., Al-qaness, M. A. A., Gandomi, A. H., & Snášel, V. (2021). Aquila Optimizer: A novel meta-heuristic optimization algorithm. Computers & Industrial Engineering, 157, 107250. https://doi.org/10.1016/j.cie.2021.107250

Aghabagherloo, A., Abadi, A., Sarkar, S., Dasu, V. A., & Preneel, B. (2025). Impact of data duplication on deep neural network-based image classifiers: Robust vs. Standard models. arxiv. https://arxiv.org/abs/2504.00638

Ahmed, M., Byreddy,S., Nutakki, A., Sikos, L. F., & Haskell-Dowland, P. (2021). ECU-IoHT: A dataset for analyzing cyberattacks in Internet of Health Things, Ad Hoc Netw 122. 102621.https://doi.org/10.1016/j.adhoc.2021.102621

Ajagbe, S.A., Mudali, P., Adigun, M. O. (2024). Internet of Things with Deep Learning Techniques for Pandemic Detection: A Comprehensive Review of Current Trends and Open Issues. Electronics,13, 2630. https://doi.org/ 10.3390/electronics13132630

Akinul Islam. J., & Mubashir, M. (2024). A comparative Analysis of Medical IoT Device Attacks Using MAchine Learning Models. Malaysian Journal of Science and Advanced Technology. https://doi.org/10.56532/mjsat.v4i4.318.

Alrefaei, A.& Ilyas, M.(2024). Using Machine Learning Multiclass Classification Technique to Detect IoT Attacks in Real Time. Sensors, 24, 4516. https://doi.org/10.3390/ s24144516

Bhushan, B., Kumar, A., Agarwal, A. K., Kumar, A., Bhattacharya, P., & Kumar, A. (2023). Towards a Secure and Sustainable Internet of Medical Things (IoMT): Requirements, Design Challenges, Security Techniques, and Future Trends. Sustainability, 15, 6177.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Dadkhah, S., Neto,E. C. P, Ferreira,R.,Molokwu,R. C., Sadeghi,S., & Ghorbani, A. (2024). CICIoMT2024: Attack vectors in healthcare devices— A multi-protocol dataset for assessing IoMT Device Security. Accessed: March, 5, 2024. doi: 10.20944/preprints202402.0898.v1.

Devi, D., Biswas, S. & Purkayastha, B. (2021). A review on solution to class imbalance problem: Undersampling approaches

Garcia, V., Sanchez, J. S., & Serrano, J. A. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Information Sciences, 505. 32-64. http://doi.org/10.1016/j.ins.2019.07.070

Gopinath M. & Sethuraman, S. C. (2023). A comprehensive survey on deep learning based malware detection techniques. Comput. Sci. Rev,47 https://doi.org/10.1016/j.cosrev.2022.100529

Hady, A. Ghubaish,A., Salman,T., Unal , D. & Jain, R. (2020). Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study, in IEEE Access, 8,. 106576-106584, 2020, doi:10.1109/ACCESS.2020.3000421.https://ieeexplore.ieee.org/document/9109651

Han, J., Pei, J.,& Kamber, M.(2011). Data mining: concepts and techniques. Amsterdam: Elsevier.

He, D., Ye, R., Chan, S., Guizani, M., & Xu, Y. (2018). Privacy in the Internet of things for smart healthcare. IEEE Communications Magazine 56(4):38–44 DOI 10.1109/MCOM.2018.1700809

Hussain, F. IoT Healthcare Security Dataset. (2023). Available online:https://www.kaggle.com/datasets/faisalmalik/iothealthcare-security-dataset (accessed on 13thMarch, 2025

Jony, A.I., Arnob, A. K. B. (2024). A long short-term memory based approach for detecting cyber attacks in IoT using CIC-IoT 2023 dataset. J. Edge comput. 3, 28-42.

Khan, M. M., & Alkhathami, M. (2024). Anomaly detection in IoT-based healthcare: machine learning for enhanced security. Scientific reports by nature portfolio. 1415872. https://doi.org/10.1038/s41598-024-56126-x

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets : A review. Science, 30(1), 25–36. https://doi.org/10.1007/978-0-387-09823-4_45

Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2021). Deduplicating Training Data MAkes Language Models Better. arXiv. Removing duplicates leads to more reliable evaluation. https://arxiv.org/abs/2107.06499

Lipsa, S.,Dash, R. K. & IvkovićNikola .(2025).An interpretable dimensional reduction technique with an explainable model for detecting attacks in Internet of Medical Things devices , Scientific Reports ,15:8718,https://doi.org/10.1038/s41598-025-93404,

Narang, P., Reddy, J. M., & Hota, C. (2013). Feature selection for detection of peer-to-peer botnet traffic. Compute 2013 - 6th ACM India Computing Convention: Next Generation Computing Paradigms and Technologies. https://doi.org/10.1145/2522548.2523133

Oyelakin, A. M., Alimi, O. M., & Abdulrauf., T. (2020). A comparative Analysis of Machine Learning Algorithms for Detecting Phising. Urls, Journal of Computer Science and Control Systems, Oredia University, Romania, 13(2):16-19. https://electroinf.uoradea.ro/index.php/jcscs/12-cercetare/reviste/jcscs/213-1st-issue-vol-13-nr-2.html.

Oyelakin, A. M. & Jimoh, R. G. (2020). Towards Building an Improved Botnet Detection Model in Highly Imbalance Botnet Dataset-A Methodological Framework. Middle East Journal of Applied Science & Technology (MEJAST) (Peer Reviewed International Journal) 3, (1), 33-38,ISSN (Online): 2582- 0974 Website: www.mejast.com

Oyelakin, A.M. & Jimoh, R.G. (2021). A Survey of Feature Extraction and Feature Selection Techniques Used in Machine Learning-Based Botnet Detection Schemes, VAWKUM Transactions on Computer Sciences, 9,1-7, available at https://vfast.org/journals/index.php/VTCS/article/view/604/658

Oyelakin, A. M., Ameen A.O., Ogundele, T.S., Salau-Ibrahim, T., Abdulrauf, U.T., Olufadi, H.I., Ajiboye, I.K., Muhammad-Thani, S. & Adeniji, I. A.(2023). Overview and Exploratory Analyses of CICIDS 2017 Intrusion Detection Dataset. Journal of Systems Engineering and Information Technology (JOSEIT)02,02, 45-52. doi: https://doi.org/10.29207/joseit.v2i2.5411

Pektas, A. & Acarman T. (2017). Effective Feature Selection for Botnet Detection Based on Network Flow Analysis, International Conference Automatics and Informatics.

Pektaş, A., & Acarman, T. (2018). Botnet detection based on network flow summary and deep learning. International Journal of Network Management, 28(6),1–15, https://doi.org/10.1002/nem.2039

Pelekoudas-Oikonomou, F., Zachos,G., Papaioannou,M., de Ree, M., Ribeiro,J. C., Mantas, G., & Rodriguez, J. (2022). Blockchain-based security mechanisms for IoMT edge networks in IoMT-based healthcare monitoring systems. Sensors, 22, no. 7, 2449, doi: 10.3390/s22072449.

Pudjihartono, N., Lee, J., & Cho, S. (2022). A review of feature selection methods for machine learning-based disease risk prediction. Frontiers in Bioinformatics, 2, 927312. https://doi.org/10.3389/fbinf.2022.927312

Pradyumna, G. R., Hegde, R.B., Bommegowda, K. B.,Tony Jan, &Naik, G, R.. (2024). Empowering Healthcare with Iomt: Evolution, Machine Learning Integration, Security, AndInteroperability Challenges. IEEE Access. doi: 10.1109/ACCESS.2024.3362239

Saka B. A., Oyelakin A. M. & Adeleke A. O. (2025). A Survey on Traditional and Deep Learning Techniques for Cyber Attack Detection in Internet of Medical Things, 2025 19th NCS International Conference on Technology and Computing tagged,Intelligent, Secure and Sustainable Innovations for Connected World held in Kano, Kano State, Nigeria

Swamynathan, M. (2017). Mastering Machine Learning with Python in Six steps, A Practical Implementation Guide to Predictive Data Analytics Using Python, DOI:10.1007/978-1-4842-2866-1_3, orhttps://tanthiamhuat.files.wordpress.com/2018/04/mastering-machine-learning-with-python-in-six-steps.pdf DOI: https://doi.org/10.29207/joseit.v2i2.5411

Yen, S.J.,& Lee, Y.S. (2009). Cluster-based under-sampling approaches for imbalanced data distributions. Expert Syst. Appl. 36, 5718–5727. https://doi.org/10.1016/j.eswa.2008.06.108.