A Review on Intrusion Detection Models in Internet of Medical Things (IoMT)

Aljorey Alqahtani, Monir Abdullah
Computing and Information Technology, University of Bisha, Bisha, 61922, P.O. Box 551, Saudi Arabia

***Abstract*—The Internet of Medical Things (IoMT) environment is highly sensitive due to the nature of medical data and its direct connection to patient health, making it a prime target for sophisticated cyberattacks. This study explores the key security challenges within IoMT, discusses how Machine Learning (ML) can enhance threat detection capabilities, and shows how XAI contributes to improving transparency and understanding of model decisions, thereby increasing trust in these systems. It reviews recent advancements in Intrusion Detection Systems (IDS) specifically designed for IoMT networks, with a focus on integrating Explainable Artificial Intelligence (XAI) and ML models. Furthermore, the study compares various algorithms and models, identifying research gaps and discussing different datasets and feature extraction techniques used for optimizing the features. The reported performance and efficiency improvements are derived from prior studies using different dataset sizes, data-splitting strategies, and feature-selection methods.**

***Keywords*—Internet of Medical Things (IoMT); IDS; Explainable Artificial Intelligence (XAI)**

I. Introduction

IoT is an innovative idea that links physical objects, devices, and everyday items to the internet to enable them to collect data by themselves, share it, and even analyze it. This technology encompasses a wide array of innovations that embed sensors, actuators, software, and network connectivity into various kinds of objects from household appliances and manufacturing equipment to automobiles, buildings, and wearable devices. Essentially, IoT provides real-time, data-driven insights and remote monitoring of such objects, rendering the physical world and digital world inextricably linked. IoT enables objects to perceive and inspect their surroundings, make decisions therefrom, and react autonomously, at times even without external intervention. This connectivity builds an ecosystem where devices collaborate to function more efficiently, accurately, and automatically, leading to smarter decision-making.

The ability for physical objects to communicate with each other and centralized systems is revolutionizing industries, driving innovation, and improving various sectors' operations. Although IoT has already shown significant promise through early applications, the technology is still in its nascent stages. As IoT technologies progress, the integration of smart sensors and connected devices into everyday life will only increase, intensifying the need for solutions that address issues like security, interoperability, protocols, and standardization. IoT devices require a reliable and scalable infrastructure to function, which presents challenges such as data management, network latency, and energy consumption. The IoTs are a revolutionary force that connects the physical and digital worlds, facilitating smarter, more responsive systems. These systems can lead to increased efficiency, reduced operational costs, and enhanced capabilities in every area of society. Despite existing challenges, the future of IoT is filled with even more promises, ushering in limitless possibilities to create connected spaces that will transform how we live, work, and interact.

IoT solutions are already gaining significant traction in verticals like healthcare, transport, manufacturing, automotive, and others, as illustrated in Fig. 1, and in healthcare, for example, IoT-enabled devices such as wearable health monitors and remote patient monitoring systems deliver valuable real-time data, improving patient care, reducing healthcare expenses, and generating better outcomes. This type of IoT is known as the IoMTs [1].

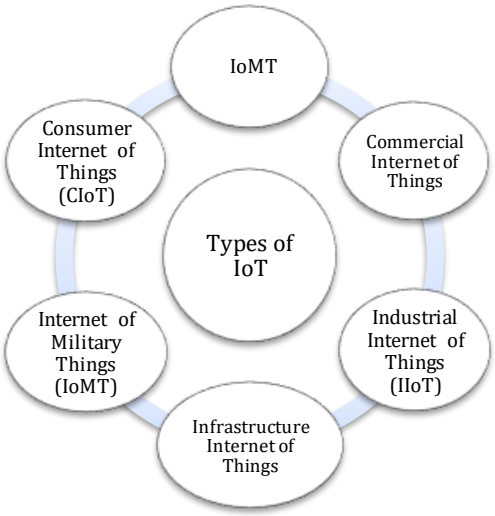

Fig. 1. IoT types.

A. The Internet of Medical Things (IoMT)

The IoMT is an intersection of medical devices with IoT. The IoMT is the future of healthcare systems, where all medical devices are interconnected and monitored on the Internet by healthcare professionals. As the IoMT evolves, it offers significant benefits, including lower healthcare costs and improved patient outcomes. The IoMT serves as a key enabler for healthcare transformation. It offers new services such as remote monitoring, senior support, and e-visits, which enhance convenience and health outcomes for patients. At the same time, these services reduce per-patient costs for healthcare organizations. However, the emergence of mobile, wearable,

and telemedicine technologies has shifted security concerns beyond traditional physical boundaries.

The increasing use of personal devices, such as smartphones and tablets, by patients and healthcare professionals to access, share, and process medical data adds complexity to the security landscape. As medical IoT devices become increasingly prevalent and process valuable data, they have become attractive targets for cyberattacks, including ransomware. However, many medical device users are unaware of the vulnerabilities present in these devices and often invest little in securing their IoT devices. However, implementing appropriate and well-known security measures can significantly mitigate several issues, such as cyber threats, weak encryption, unauthorized access, design flaws, insecure networks, lack of security awareness, compatibility issues, update management, dependency on the cloud, and lack of uniform standards. Therefore, security measures, such as IDS, can significantly reduce the risks associated with exploiting these vulnerabilities [2].

B. Intrusion Detection System (IDS)

ID refers to the organized process of monitoring and analyzing activities within a computer system or network to spot signs of unauthorized access, security breaches, or attempts to compromise the system's confidentiality, integrity, or availability, as well as efforts to bypass its security defenses. Such intrusions may occur when external parties gain unauthorized access through the Internet, when legitimate users attempt to escalate their privileges beyond what is allowed, or when they misuse the access they already have. To support this process, IDSs are employed—these can be either software applications or dedicated hardware devices designed to automate the monitoring and analysis of suspicious activities.

ID plays a vital role in helping organizations protect their systems from the increasing risks that accompany increased connectivity and reliance on digital infrastructure. Given the breadth and complexity of today's cyber threats, the question for security teams is not whether to implement ID, but rather to focus on choosing the right features and functionality for their needs. IDS are now an essential part of any organization's security program. There are several good reasons to deploy them:

- Deterrence: By increasing the chances of detecting and punishing attackers, IDSs discourage malicious behavior and reduce the risk of system misuse.
- Attack detection: IDSs are essential for detecting attacks and vulnerabilities that other security measures cannot prevent.
- Attack precursor detection: IDs can identify early signs of potential attacks, allowing organizations to respond proactively.
- Threat documentation: IDSs provide an organization with a record of ongoing threats, helping to understand the current security landscape.
- Security quality monitoring: IDSs serve as a tool for evaluating and managing security measures, especially in large and complex environments.
- Improved recovery and analysis: IDSs provide valuable information about breaches, making it easier to diagnose, recover from, and correct underlying issues.

By increasing the expected risk of detection and punishment, IDS acts as a powerful deterrent against attackers [3].

II. Related Works

Over the past few years, there have been numerous studies aimed at creating and improving IDS for IoT and IoMT networks. The studies aim to utilize novel methods to enhance the capacity of the systems to detect and manage threats effectively. The research in [4] explores the detection of cyberattacks in IoT networks through a deep learning (DL)-based IDS. The system employs an autoencoder (AE) model that optimizes detection time and offers enhanced precision with 99.76% detection. The AE model operates through encoding and decoding network packets to identify malicious activities. The experiment used the UNSW-NB15 dataset. This research in [5] explores ML algorithms to maximize IDS performance across IoT networks. With the ToN-IoT dataset, the authors implemented Random Forest (RF) and XGBoost, both ensemble learning methods that were very efficient in ID. The solution described reached 99% accuracy.

The study in [6] explores the integration of blockchain technology with anomaly-based ID systems to enhance the (IoT) networks security. This study leverages Convolutional Neural Networks (CNNs) and Artificial NNs (ANNs). They are trained on the IoT-23 dataset, achieving an impressive accuracy of 99.8%. This approach not only enhances security but also supports scalability within IoT environments. In [7], the researchers present an IDS based on federated transfer learning to secure IoMT devices. Their method uses a Deep NN (DNN) to train models locally on the endpoints, while preserving user data privacy. Testing using the CICIDS2017 dataset yielded a high accuracy rate of 99.51%, confirming the system's effectiveness in protecting sensitive healthcare information.

The work presented in [8] focuses on building an IDS specifically designed for IoMT applications using the CICIoMT-2024 dataset. The study evaluates the performance of three boosting algorithms—XGBoost, AdaBoost, and CatBoost—for classifying network traffic. Among these algorithms, XGBoost achieved the best results with an accuracy of 95.01%, followed by AdaBoost with 92.89% and CatBoost with 88.37%. Additionally, the researchers applied XAI techniques to make the models more interpretable.

The study in [9] aims to combat Distributed Denial of Service (DDoS) attacks in blockchain-enabled IoMT networks using ML-based IDs. The study compares the performance of XGBoost, Decision Tree (DT), and RF models using the CICIoMT-2024 dataset. The results demonstrate the system's strong potential to effectively detect and mitigate DDoS attacks, achieving an accuracy of 99.99%. In [10], an anomaly-based IoT network IDS using a DNN with filter-based feature selection to exclude highly correlated features is proposed. The model started by scoring 84% accuracy on UNSW-NB15. To address class imbalance, the authors utilized Generative Adversarial Networks (GANs) to produce synthetic data for minority attack classes to boost accuracy to 91%. The work in [11] suggests an

EML algorithm-based IoT network IDS using the Message Queuing Telemetry Transport (MQTT) protocol.

The authors created the SEN-MQTTSET dataset from three experiment scenarios: normal traffic, subscriber attacks, and broker attacks. Feature extraction was performed using an ensemble statistical algorithm, and algorithms such as Random Forest (RF) and support vector machine (SVM) were used. The system achieved a 99% accuracy level as well as network latency measurement for normal and malicious traffic. The research work in [12] discusses a novel attacker detection method for an edge-focused IoMT structure based on the Swarm NN (Swarm-NN) framework. The system, implemented on the ToN-IoT dataset as well as a real-time secured dataset, achieved 99.5% accuracy. The study in [13] focuses on monitoring network traffic and medical devices to detect malicious and unintended activity. To improve performance while reducing computational costs, the study uses mutual information (MI) and XGBoost as filter-based feature selection techniques. Evaluating the system on the CICIDS2017 dataset, it achieved a high accuracy of 98.79%. In [14], the authors propose *SafetyMed*, an intelligent IDS that combines long-short-term memory (LSTM) and CNNs networks. This hybrid model is designed to analyze sequential and network data to detect intrusion attempts. After validation on the CICIDS2017 dataset, *SafetyMed* demonstrated an accuracy of 97.63%.

The study in [15] presents a swarm neural network-based ID approach specifically designed for data-intensive IoMT systems. The model was tested on the real-time NF-ToN-IoT dataset and achieved an accuracy of 89.0%, demonstrating its practical applicability in healthcare settings. The work in [16] presents a Software-as-a-Service (SaaS)-based IDS specifically developed to address the unique security challenges of IoMT systems. This approach incorporates PSO for feature engineering and was evaluated on the WUSTL-EHMS-2020 dataset, achieving an accuracy of 96.56%. The comparison of the previous studies is shown in Table I.

TABLE I. COMPARISON OF PREVIOUS STUDIES

Ref	Problem	Methods	Dataset	Accuracy
[4]	Detecting malicious activity on IoT network traffic and improving ID in 5G networks	Autoencoder-based DL technique with DNN, XGBoost, Adaboost, ExtraTreeClassifier, RF	UNSW-NB15	99.76%
[5]	The increasing security threats to IoT networks due to the rapid expansion of connected devices	ML-based classification models, including Ridge Classifier, XGBoost, Logistic Regression, RF, and Gradient Boosting	ToN-IoT dataset	99.85%
[6]	Increasing security challenges in IoT networks due to growing threats and scalability issues	A combination of Blockchain for access control and Anomaly-Based IDS using lightweight ML models on FPGA hardware accelerators	IoT-23 dataset	94%
[7]	Increasing cyberattacks on IoMT devices and the need for data privacy	Federated Transfer Learning using DNN	CICIDS2017	99.51%
[8]	cybersecurity threats to IoMT devices and the need for explainable AI-based IDS	Boosting ensemble methods (XGBoost, AdaBoost, CatBoost) with XAI	CICIoMT-2024	XGBoost: 95.01%,
[9]	Vulnerability of Blockchain-enabled IoMT networks to DDoS attacks, affecting network availability and security	ML-based IDS using XGBoost, DT and RF	CICIoMT2024	99.99%
[10]	Traditional threat detection methods struggle to detect new threats and deal with imbalanced data sets	DNN with filter-based feature selection and Generative Adversarial Networks to generate synthetic data for minority attack category.	UNSW-NB15 dataset	84% without GANs, 91% with GANs
[11]	IoT networks face increased cyber-attacks and time constraints in communication.	Elite ML (EML) algorithms with an ensemble statistical multi-view cascade feature generation.	SEN-MQTTSET dataset	99%
[12]	IoMT networks are vulnerable to cyber-attacks during data transmission, leading to privacy leakage and security risks.	Swarm-NN (Swarm-NN) with an Empirical Intelligent Agent (EIA) for ID and health data analysis.	ToN-IoT dataset	99.5%
[13]	IoMT networks are vulnerable to cyberattacks, necessitating effective and accurate IDS.	Tree-based ML (DT, RF, XGBoost, CatBoost) with Mutual Information and XGBoost Feature Selection (MI-XGBoost).	CICIDS2017 dataset	98.79%
[14]	IoMT devices are vulnerable to cyberattacks due to limited computational power, simplified architecture, making them easy targets	Hybrid CNN and LSTM networks for ID from sequential and grid data.	CIC-IDS2017 dataset	97.63%
[15]	IoMT devices have limited storage capacity, computing power, which requires data transfer to external systems, which then leads to security vulnerabilities.	Swarm-NN model for detecting intruders during data transfer in IoMT systems.	NF-ToN-IoT dataset	89.0%
[16]	IoMT devices are resource-constrained and vulnerable to cyberattacks.	SaaS-based IDS with PSO, ML/DL models, and SHAP.	WUSTL-EHMS-2020	96.57%

III. ML APPROACHES

ML models use some algorithms to learn and identify complex patterns in data. ML is all about making computers capable of identifying patterns in data and making decisions or predictions based on the intelligence gathered. There are several primary approaches to ML:

- Supervised Learning: In this approach, the model learns from labelled data, where both the inputs and outputs are provided. The goal is to make accurate predictions or classifications, such as identifying diseases based on medical data or predicting stock prices. Examples include DTs, Linear Regression (LR), and SVM.

- Unsupervised Learning: Unsupervised learning addresses data without pre-defined labels. The model tries to identify inherent structures or patterns, such as clustering similar data points or reducing the dimension of complex datasets. Common methods are k-means clustering, principal component analysis (PCA), and autoencoders.
- Reinforcement Learning (RL): RL is a learning paradigm that mimics human trial-and-error learning. An agent takes an action in an environment, receiving feedback (reward or penalty) based on the outcome of the action taken. RL has wide-ranging applications in robotics, game playing (e.g., AlphaGo), and autonomous cars.
- Semi-supervised and Self-supervised Learning: These methods are between supervised learning and unsupervised learning, using both labelled and unlabelled data or extracting features from the data itself to reduce the amount of labelled data. With continuous advancements, ML is becoming more powerful, adaptive, and integrated into everyday technologies [17].

A. Classification

Classification is a supervised ML in which the aim is to identify a categorical class label for the input. When classifying, the model has been trained using labelled data so that every one of the inputs corresponds to a class label. After training, new, unseen data can be employed to classify, based on patterns learned in the course of training. Classification operations are common in various disciplines like health, finance, and information security [18]. There are two types of classification:

1) Binary: It is a supervised learning problem in which the goal is to classify input data into one of the two potential classes. The model is trained on labelled data to distinguish between two outcomes, most commonly being 0 or 1, True or False, or Positive or Negative. It is widely applied in all domains like spam detection, disease diagnosis, fraud detection, and opinion extraction. In [19], the authors describe the various simple performance measures used in binary class assessment as belonging to three main families: the measures depending on one classification threshold (raw and composite measures), the measures depending on the probabilistic interpretation of error, and the classification measures. The authors also cover graphical methods such as ROC curves and precision–recall curves and outline statistical methods for examining the significance of performance measures and calculating confidence intervals. They provide a simplified example to illustrate the calculation of these measures and the interrelationships between them, emphasizing the importance of choosing appropriate performance measures based on specific classification objectives, especially in the context of class imbalance. In [20], the author provides a comprehensive review of binary performance assessment, focusing on the analysis, relationships, and classification of different performance tools. The study classifies these tools into "performance measures" and "performance metrics", and defines them semantically and formally to clarify their use across different domains. Several new concepts, such as binary, complementarity, and normalization, are introduced to examine the similarities, redundancies, and dependencies between different tools. The main contributions of the study are the Periodic Table of Performance Tools (PToPI), a visual representation that organizes performance tools systematically, helping researchers choose appropriate metrics.

2) Multi-class: It is a statistical and ML classification task that involves more than two classes. It involves classifying data into one of the possible classes. Each sample can only be classified into one class. In [21], the author provides a detailed review of the evaluation metrics used in multiclass classification, emphasizing their role in evaluating ML models. The discussion begins with basic concepts such as precision, recall, and the confusion matrix, which serve as the basis for more advanced metrics. The study explains how precision, despite its widespread use, can be misleading, especially in imbalanced datasets, leading to the introduction of balanced precision and weighted variances to address this issue. Cross-entropy is introduced as a metric that assesses the divergence between predicted and actual probability distributions. In addition, the author examines more advanced evaluation techniques, such as Matthew's Correlation Coefficient (MCC), which considers all components of the confusion matrix, and Cohen's kappa, which measures the concordance of predicted versus actual classes with chance adjustment. The talk emphasizes the importance of applying the right evaluation methods based on the specific problems of the dataset, such as class imbalance or the need for good prediction in certain classes.

B. ML Key Models

DT models are a popular and easy-to-understand ML method, applied in both classification and regression problems. The method works by splitting the dataset into subsets depending on feature values, allowing decisions to culminate in classification or prediction outcomes. According to authors in [22], DTs encompass a number of algorithms, including ID3, C4.5, CART, CHAID, and QUEST. These algorithms have been widely used in various fields, including medical diagnosis, text classification, and image processing, because of their interpretability and simplicity.

SVM models are powerful supervised learning algorithms well-suited for both classification and regression tasks, with a particular strength in handling classification problems. The basic idea behind SVM is to find the optimal level that separates the data points into different categories within an N-dimensional space. By maximizing the margin between the closest data points of each class, SVMs enhance both the accuracy and generalization of the model. According to the findings in [23], SVMs can efficiently process large-scale datasets, making them highly effective in terms of computational performance and accuracy. The study also explored various optimization strategies, such as integrating fuzzy membership functions into multi-kernel learning frameworks, and identifying optimal solution spaces based on dataset size, further demonstrating

SVM's scalability and adaptability to complex data environments.

The k-Nearest Neighbors (k-NN) model is a straightforward, yet widely used supervised learning method, especially effective for classification and regression tasks. As a non-parametric approach, it works by assigning a class to a new data point based on the majority class among its closest neighbors. Its simplicity and effectiveness make it a popular choice in many ML applications. In [24], the k-NN algorithm was used to categorize data into positive and negative classes. The results showed high accuracy, confirming the algorithm's strength in reliable and accurate classification tasks.

Naive Bayes (NBs) model is a probabilistic classifier in the field of ML that is commonly applied in classification, particularly where we have large data. It functions on the principles of Bayes' theorem, which presumes features to be conditionally independent, i.e., the presence of one feature does not influence the presence of another. The algorithm is specifically utilized in domains such as text classification, spam filtering, and sentiment analysis because of its simplicity and efficiency. Naïve Bayes is known for its speed and effectiveness on high-dimensional data, and it serves as a solid foundation for various classification problems. In [25], the researchers used the Naïve Bayes algorithm. Two different data distribution techniques, Hold-Out and 10-Cross Fold Validation, were used to evaluate the algorithm's performance. It was found that using the Naïve Bayes algorithm with the Hold-Out method achieves better accuracy.

RFs model is a popular ML model that builds a collection of DTs and merges their individual predictions to generate the final result. The approach is suitable for both classification (prediction of classes) and regression (prediction of continuous outcomes) problems. It works by randomly selecting subsets of data and features to train multiple DTs, which helps improve accuracy and prevent overfitting. The overall outcome is based on the majority vote (in classification) or average (in regression) of the trees, making the model more stable and reliable than a single decision tree. Its strength lies in its ability to handle complex datasets and adapt to a variety of patterns, making it a robust and accurate algorithm, as in [26]. The author conducted a novel spatial RFs technique that enhances traditional RFs by incorporating higher-order spatial statistics. This approach utilizes local spatial-spectral information to effectively identify intrinsic heterogeneity, spatial interactions, and sophisticated spatial patterns, hence presenting a more sophisticated framework for geoscience data modeling and analysis. Further, it highlights the superiority of this approach compared to traditional RFs, specifically in their potential to generate spatially coherent predictions.

Neural Networks (NN) models: A neural network is a computer system modeled on the biological neural networks found in the human brain. It is composed of interconnected neurons, or nodes, that work together to recognize patterns, make decisions, and solve complex problems. In [27], the authors discussed various optimization methods which are employed to enhance artificial NNs (ANNs) through the use of parameter optimization methods including network architecture, hidden neuron numbers, and learning rates. Several algorithms include the particle swarm optimization (PSO), Genetic Algorithm (GA), backtracking search algorithm (BSA), Artificial Bee Colony (ABC), among others, help to maximize both the effectiveness and efficacy of ANNs.

IV. Datasets in IoMT

In IoMT, datasets play a pivotal role in developing and evaluating ML models for healthcare applications, as shown in Table II. This data is typically collected from various IoT devices, some from different medical devices. Common datasets used in the IoMT include:

UNSW-NB15 dataset is a premier benchmark to assess network IDS (NIDS). It was developed by the Australian Centre for Cyber Security (ACCS) at UNSW Canberra with the vision to overcome the drawbacks of existing datasets by incorporating contemporary attack scenarios and actual network traffic. The dataset consists of raw network packet captures processed using Bro-IDS and Argus tools, resulting in 49 features that include flow-based, basic connection characteristics, and content-based characteristics. It contains both normal and malicious traffic, and covers a variety of attack types such as DoS, exploits, fuzzers, backdoors, and reconnaissance. UNSW-NB15 is widely used in cybersecurity research, particularly in ID, anomaly detection, and ML-based security applications. Its richness in attack diversity and realistic traffic patterns makes it a valuable resource for testing the effectiveness of modern NIDS solutions [28].

ToN-IoT (Telemetry and Network of IoT) dataset is a modern cybersecurity dataset designed for ID in Internet of Things (IoT) and Industrial IoT (IIoT) environments. It was developed by the UNSW Canberra Cyber Range Lab and includes telemetry data from various IoT devices, network traffic, and system logs, so it is a large set for cybersecurity research. The dataset includes both normal and abnormal and has a broad variety of cyber threats such as denial of service (DoS), ransomware, backdoors, and reconnaissance. It embraces multiple data types, including IoT sensor logs, operating systems (Windows and Linux), and network traffic, to facilitate cross-layer security analysis. ToN-IoT is widely used in ML-based ID and anomaly detection due to its realistic attack scenarios and multi-source data collection, making it highly relevant to modern IoT security problems [29].

IoT-23 dataset is a public database, and it is intended for educational research purposes in the area of malware and network ID in IoT settings. IoT-23 was created by Stratosphere Lab in collaboration with Avast, and contains 23 different scenarios, including normal IoT network traffic and malicious traffic generated by different malware families. The dataset consists of classified network traffic captures (PCAP files) along with extracted flow-based features, allowing researchers to analyze various attack behaviors such as a DDoS, botnets, and control (C&C) communications. IoT-23 is widely used in cybersecurity research to develop and evaluate ML models aimed at detecting and mitigating IoT-based cyber threats [30].

The CICIDS2017 dataset is generated by the Canadian Institute for Cybersecurity (CIC), is a premier benchmark for the testing of IDS and conducting cybersecurity research. The dataset mimics actual network traffic by including both

legitimate activity and a variety of attack types, such as DDoS attacks, brute force attacks, botnet traffic, intrusions, and web attacks. It consists of detailed network flow data captured over a five-day period, including packet captures (PCAPs) and extracted flow-based features. CICIDS2017 is valuable for training ML models due to its inclusion of time-stamped traffic, classified attack categories, and real-world user behavioral profiles. The dataset is widely used in anomaly detection, performance evaluation of IDS, and cybersecurity analytics, making it a primary resource for advancing network security research [31].

The CICIoMT-2024 dataset is an extensive benchmark dataset created for the assessment of security controls within the environment of the IoMT. Created by the Canadian Institute for Cybersecurity, the dataset comprises network traffic data collected from an IoMT testbed composed of 40 devices, of which 25 were real devices and 15 simulated devices, with the perspective of representing the different protocols utilized in healthcare settings like Wi-Fi, MQTT, and Bluetooth. In order to replicate the real-world threat scenarios, 18 different attacks have been performed on this testbed, which are divided into five main categories: DDoS, DoS, reconnaissance, MQTT-based attacks, and spoofing. This dataset serves as a useful tool for researchers and practitioners interested in designing and testing IDSs and other security mechanisms for IoMT infrastructure [32].

The SEN-MQTTSET dataset is a specialized resource developed to enhance IDSs within IoT environments using the MQTT protocol. This dataset includes three distinct scenarios: normal operations, attacks targeting subscribers, and attacks on intermediaries. To extract meaningful features from the raw data, a statistics-based multi-view sequential feature generation algorithm was used, resulting in a multi-context feature set [33].

The NF-ToN-IoT dataset is the NetFlow-based version of the ToN-IoT dataset to enhance research on network ID in IoT networks. By transforming ToN-IoT raw packet captures (pcaps) into NetFlow logs, NF-ToN-IoT gives us the flow-based view of the network traffic and therefore, it is valuable to investigate communication patterns as well as for anomaly detection. The dataset contains a total of about 16.9 million data flows, where 63.99% are marked as attack samples and 36.01% are labeled as benign. The attack categories are Backdoor, DoS, DDoS, Injection, Man-in-the-Middle (MITM), Password, Ransomware, Scanning, and XSS. Each flow is described using 12 NetFlow features, including source and destination IP addresses, ports, protocols, and byte counts. NF-ToN-IoT is publicly available and is a valuable resource for training and testing ML models for enhancing the security of IoT networks [34].

The WUSTL-EHMS-2020 dataset is a targeted resource aimed at promoting cybersecurity research in the IoMT context. The dataset was obtained using the Enhanced Healthcare Monitoring System (EHMS) Real-Time Testbed located at Washington University in St. Louis, and it uniquely integrates network flow measurements with patient biometric information, thereby filling the gap of datasets that bring these two elements together. It has 16,318 samples, of which 14,272 are labeled as normal, and 2,046 are attack instances, and has 44 distinct features—35 for network metrics and 8 for patient vital signs. The dataset includes man-in-the-middle attacks, such as spoofing and data injection, and data confidentiality and integrity violations. Researchers are using WUSTL-EHMS-2020 to develop and evaluate IDSs specifically designed for IoMT environments, leveraging ML techniques to analyze the complex interaction between network activity and medical data [35].

The dataset is typically divided into training, validation, and test sets for model development and evaluation. Common split ratios include 70% for training, 15% for validation, and 15% for testing, or 80% for training and 20% for testing, to ensure fair evaluation and good generalization performance.

TABLE II. IOMT DATASETS

Ref.	Dataset	Field	Year	Total Features	Target attack
[28]	UNSW-NB15	Cybersecurity, Network ID	2015	49	DoS, Fuzzers, Backdoor, Exploits, Generic, Reconnaissance, Shellcode, Worms
[29]	ToN-IoT	IoT network traffic.	2020	44	DDoS, Mirai, Keylogging, Scanning, DoS, Backdoor, Ransomware, Injection, XSS, Password attack, MITM
[30]	IoT-23	IoT network traffic.	2020	47	DDoS, DoS, Brute Force, Port Scan, MITM, Malware, Backdoor, Worm, Credential Stuffing, Botnet attacks
[31]	CICIDS2017	Network traffic data from various real-world scenarios	2017	80	DDoS, Brute Force, DoS, Port Scan, Heartbleed, Botnet, Web Attack, SSH, Infiltration, etc.
[32]	CICIoMT-2024	Network traffic data from IoMT devices	2024	44	DDoS, DoS, Spoofing, Recon, MQTT
[33]	The SEN-MQTTSET	Network traffic data from IoT devices using MQTT protocol	2021	120	DoS, Subscriber Attack, Broker Attack
[34]	NF-ToN-IoT v1, v2, v3	Network traffic data from IoT devices	2021-2025	8, 43, 53	Backdoor, DoS, DDoS, Injection, Scanning, XSS
[35]	WUSTL-EHMS-2020	IoMT network	2020	44	MITM attacks: Spoofing, Data Injection

V. EXPLAINABLE AI (XAI) ALGORITHMS

XAI refers to a group of methods and techniques designed to enhance the transparency, interpretability, and trustworthiness of AI models. Traditional AI models, especially DL and complex ML models, often function as "black boxes", making it difficult to understand how they arrive at specific decisions. XAI algorithms aim to bridge this gap by providing insights into model predictions, ensuring accountability, and facilitating human-AI collaboration. XAI algorithms can be broadly classified into two types:

A. Model-Specific XAI Algorithms

These methods are built into specific models and provide intrinsic interpretability.

- DTs: A hierarchical structure that allows for straightforward interpretation by mapping decision paths.
- LR: Provides direct interpretability through feature coefficients that indicate the impact of each variable.
- Rule-Based Models: Generate human-readable rules that explain decision-making, such as in fuzzy logic systems.

B. Post-Hoc XAI Algorithms

These techniques analyze and interpret complex, pre-trained models without altering their structure.

- Shapley Additive Explanations (SHAP): Founded on cooperative game theory, SHAP allocates a significant score to every feature, indicating its contribution to the prediction made by the model.
- Local Interpretable Model-agnostic Explanations (LIME): Creates local approximations of a black-box model by training simple surrogate models in the vicinity of single predictions.
- Gradient-weighted Class Activation Mapping (Grad-CAM): Used for DL to reveal the regions of an image that supported a model's decision.
- Counterfactual Explanations: Explain decisions by identifying minimal changes in input that would lead to a different output, providing actionable insights [36]. These are all XAI techniques, as shown in Table III:

TABLE III. COMPARISON OF XAI TECHNIQUES

Technique	Description	Strengths	Weakness	Use Cases
DT	A hierarchical structure where decisions are made based on feature values.	Transparent, easy to visualize, interpretable.	Prone to overfitting, less powerful for complex tasks.	Customer segmentation, medical diagnosis, credit scoring.
LR	Predicts outputs based on a weighted sum of input features.	Simple, easy to interpret, direct feature influence.	Limited to linear relationships, sensitive to outliers.	Price prediction, demand forecasting, stock market prediction.
Rule-Based Models	Make decisions based on predefined or learned rules.	Transparent, easy to implement, interpretable.	Can become overly complex, less powerful for complex patterns.	Expert systems, diagnostic systems, regulatory compliance.
SHAP	Assigns an important score to each feature based on game theory (Shapley values).	Theoretically grounded, consistent across models.	Computationally expensive, slow for large models.	Feature importance analysis, model interpretation, ML.
LIME	Approximates complex models with simpler interpretable models around individual predictions.	Model-agnostics provides local explanations.	Local explanations may not generalize, computationally expensive.	Image classification, text classification, fraud detection.
Grad-CAM	Visualizes which parts of an image influenced the model's decision.	Provides intuitive visualizations, works well with CNNs.	Limited to CNNs, hard to interpret for complex models.	Image classification, object detection, medical imaging.
Counterfactual Explanations	Explain a decision by identifying minimal changes in input to alter the output.	Actionable, easy to understand, intuitive.	May not be meaningful in complex models, computationally intensive.	Decision support, recommender systems, financial services.

VI. FEATURE SELECTION TECHNIQUES IN ID

Feature selection is a necessary process in developing IDSs to reach dimensionality reduction, better accuracy, and better computational efficiency. By selecting the most informative features from network traffic data, IDS can gain a higher classification accuracy while, in the meantime, decreasing the number of false positives and negatives. A few optimization algorithms in feature selection, specifically bio-inspired ones such as GA, PSO, GWO, and ABC, play an important role in this process. GA ensures solution diversity through crossover and mutation, but it can be computationally expensive.

PSO is efficient and simple, but may suffer from early convergence. GWO effectively balances exploration and exploitation, making it suitable for high-dimensional IDS datasets. ABC offers flexibility and robustness but may converge slowly than other algorithms. Despite their limitations, these methods continue to evolve, addressing weaknesses and improving feature selection performance, ultimately enhancing IDS effectiveness. The selection of an appropriate feature selection technique relies on factors like dataset size, feature complexity, and the desired balance between computational cost and classification performance. Proper feature selection can significantly enhance IDS effectiveness by improving detection rates while reducing processing overhead [37].

VII. PRACTICAL DEPLOYMENT: CONSIDERATIONS AND LIMITATIONS

Practical deployment of IoT and IoMT intrusion detection systems also reveals important evaluation limitations commonly observed in existing studies. In particular, several datasets rely on labeling mechanisms derived from device identifiers, such as source MAC addresses. When MAC-related or identity-derived attributes are simultaneously included in the feature space, learning models may unintentionally capture device identity rather than genuine attack behavior, leading to label–feature leakage and overly optimistic performance results. To mitigate this risk, robust experimental practices explicitly exclude direct device identifiers (e.g., raw MAC addresses, organizationally unique identifiers, or vendor-specific prefixes) from the feature set and enforce device-level data partitioning, ensuring that traffic generated by the same device does not appear across training and testing splits. Clear documentation of feature

inclusion/exclusion decisions and data-splitting strategies is therefore essential for realistic, generalizable, and reproducible IoMT IDS evaluations.

VIII. Conclusion

The review shows that combining ML and XAI represents a promising direction in designing IDSs for IoMT environments. ML techniques enable accurate detection of complex threats, while XAI helps interpret these decisions and support trustworthy responses in sensitive medical contexts. However, challenges remain regarding scalability, real-time data handling, and privacy protection. The study recommends further studies to develop more efficient hybrid models, increase the adoption of XAI to ensure transparency, and establish standardized testing frameworks suitable for IoMT systems. Additionally, existing evaluations may suffer from label–feature leakage when device identifiers such as MAC addresses are used for labeling while also being included in the feature space. To ensure realistic and generalizable results, future studies should exclude direct device identifiers and apply device-level data partitioning between training and testing sets.

Acknowledgment

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at the University of Bisha for the financial support through the Graduate Students Research Support Program.

References

[1] Gokhale, P., Bhat, O., & Bhat, S. (2018). Introduction to IOT. International Advanced Research Journal in Science, Engineering and Technology, 5(1), 41-44.

[2] Razdan, S., & Sharma, S. (2022). Internet of medical things (IoMT): Overview, emerging technologies, and case studies. IETE technical review, 39(4), 775-788.

[3] Bace, R. G., & Mell, P. (2001). Intrusion detection systems.

[4] Yadav, N., Pandey, S., Khambharia, A., and Gupta, D. (2022). Intrusion Detection System on IoT with 5G Network Using Deep Learning. Wireless Communications and Mobile Computing, 2022(1), 9304689.

[5] Kaddour, H., Das, S., Bajgai, R., Sanchez, A., Sanchez, J., Chiu, S. C., ... & Fouda, M. M. (2024, April). Evaluating the Performance of Machine Learning-Based Classification Models for IoT Intrusion Detection. In 2024 IEEE Opportunity Research Scholars Symposium (ORSS) (pp. 84-87). IEEE.

[6] Ngo, D. M., Lightbody, D., Temko, A., Murphy, C. C., & Popovici, E. (2024). A Scalable Security Approach in IoT Networks: Smart Contracts and Anomaly-based IDS for Gateways using Hardware Accelerators. IEEE Access.

[7] Otoum, Y., Wan, Y., & Nayak, A. (2021, December). Federated transfer learning-based ids for the internet of medical things (iomt). In 2021 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.

[8] Sohail, F., Bhatti, M. A. M., Awais, M., & Iqtidar, A. (2024, October). Explainable Boosting Ensemble Methods for Intrusion Detection in Internet of Medical Things (IoMT) Applications. In 2024 4th International Conference on Digital Futures and Transformative Technologies (ICoDT2) (pp. 1-8). IEEE.

[9] Akkal, M., Cherbal, S., Kharoubi, K., Annane, B., Gawanmeh, A., & Lakhlef, H. (2024, November). An Intrusion Detection System For Detecting DDoS Attacks In Blockchain-Enabled IoMT Networks. In 2024 7th International Conference on Signal Processing and Information Security (ICSPIS) (pp. 1-6). IEEE.

[10] Sharma, B., Sharma, L., Lal, C., & Roy, S. (2023). Anomaly based network intrusion detection for IoT attacks using deep learning technique. Computers and Electrical Engineering, 107, 108626.

[11] Siddharthan, H., Deepa, T., & Chandhar, P. (2022). Senmqtt-set: An intelligent intrusion detection in iot-mqtt networks using ensemble multi cascade features. IEEE Access, 10, 33095-33110.

[12] Nandy, S., Adhikari, M., Khan, M. A., Menon, V. G., & Verma, S. (2021). An intrusion detection mechanism for secured IoMT framework based on swarm-neural network. IEEE Journal of Biomedical and Health Informatics, 26(5), 1969-1976.

[13] Balhareth, G., & Ilyas, M. (2024). Optimized intrusion detection for IoMT networks with tree-based machine learning and filter-based feature selection. Sensors, 24(17), 5712.

[14] Faruqui, N., Yousuf, M. A., Whaiduzzaman, M., Azad, A. K. M., Alyami, S. A., Liò, P., ... & Moni, M. A. (2023). SafetyMed: a novel IoMT intrusion detection system using CNN-LSTM hybridization. Electronics, 12(17), 3541.

[15] Awotunde, J. B., Abiodun, K. M., Adeniyi, E. A., Folorunso, S. O., & Jimoh, R. G. (2021, November). A deep learning-based intrusion detection technique for a secured IoMT system. In International Conference on Informatics and Intelligent Applications (pp. 50-62). Cham: Springer International Publishing.

[16] Aljuhani, A. Alamri, P. Kumar and A. Jolfaei, "An Intelligent and Explainable SaaS-Based Intrusion Detection System for Resource-Constrained IoMT," in IEEE Internet of Things Journal, vol. 11, no. 15, pp. 25454-25463, 1 Aug.1, 2024, doi: 10.1109/JIOT.2023.3327024.

[17] Sarhan, M., Layeghy, S., Moustafa, N., Gallagher, M., & Portmann, M. (2024). Feature extraction for machine learning-based intrusion detection in IoT networks. Digital Communications and Networks, 10(1), 205-216.

[18] Shi, Y., Zhang, W., Yang, Y., Murzin, A. G., Falcon, B., Kotecha, A., ... & Scheres, S. H. (2021). Structure-based classification of tauopathies. Nature, 598(7880), 359-363.

[19] Berrar, D. (2019). Performance measures for binary classification.

[20] Canbek, G., Taskaya Temizel, T., & Sagiroglu, S. (2022). PToPI: A comprehensive review, analysis, and knowledge representation of binary classification performance measures/metrics. SN Computer Science, 4(1), 13.

[21] Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

[22] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of applied science and technology trends, 2(01), 20-28.

[23] Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of support vector machine algorithm in big data background. Mathematical Problems in Engineering, 2021(1), 5594899.

[24] Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) algorithm for public sentiment analysis of online learning. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 15(2), 121-130.

[25] Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store: The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews On Google Play Store. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 2(1), 47-54.

[26] Talebi, H., Peeters, L. J., Otto, A., & Tolosana-Delgado, R. (2022). A truly spatial random forests algorithm for geoscience data analysis and modelling. Mathematical Geosciences, 54(1), 1-22.

[27] Abdolrasol, M. G., Hussain, S. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., ... & Milad, A. (2021). Artificial neural networks based optimization techniques: A review. Electronics, 10(21), 2689.

[28] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In Proceedings of Military Communications and Information Systems Conference.

[29] Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., & Anwar, A. (2020). TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems. Ieee Access, 8, 165130-165150.

[30] Abdalgawad, N., Sajun, A., Kaddoura, Y., Zualkernan, I. A., & Aloul, F. (2021). Generative deep learning to detect cyberattacks for the IoT-23 dataset. IEEE Access, 10, 6430-6441.

[31] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 1(2018), 108-116.

[32] Dadkhah, S., Neto, E. C. P., Ferreira, R., Molokwu, R. C., Sadeghi, S., & Ghorbani, A. A. (2024). CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT. Internet of Things, 28, 101351.

[33] Vaccari, I., Chiola, G., Aiello, M., Mongelli, M., & Cambiaso, E. (2020). MQTTset, a new dataset for machine learning techniques on MQTT. Sensors, 20(22), 6578.

[34] Mancilla, R. O., ENG, F. C., & Diaz, J. M. V. (2020). D4. 5 Intrusion detection for IoT-based context and networks.

[35] Ravi, V., Pham, T. D., & Alazab, M. (2023). Deep learning-based network intrusion detection system for Internet of medical things. IEEE internet of things magazine, 6(2), 50-54.

[36] Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., ... & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. ACM Computing Surveys, 55(9), 1-33.

[37] Selvarajan, S. (2024). A comprehensive study on modern optimization techniques for engineering applications. Artificial Intelligence Review, 57(8), 194.