Results in Engineering 23 (2024) 102659

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

An RFE/Ridge-ML/DL based anomaly intrusion detection approach for securing IoMT system

Ghita Lazrek *, Kaouthar Chetioui, Younes Balboul, Said Mazer, Moulhime El bekkali

Lab LIASSE, ENSA, Sidi Mohamed Ben Abdellah University, Fez, Morocco

A R T I C L E I N F O

Keywords:
Anomaly intrusion detection system (AIDS)
Cyber-attack
Deep learning (DL)
Internet of medical things (IoMT)
Machine learning (ML)
Security

A B S T R A C T

Smart healthcare is one of the promising areas of the Internet of Things (IoT), particularly in the case of the Covid-19 pandemic. Real-time patient monitoring and remote diagnostics facilitate better medical services to preserve human lives using Internet of Medical Things (IoMT) technology. Regardless of the numerous benefits, IoMT devices are susceptible to sophisticated cyber-attacks at a breakneck pace, which lead to tampering with healthcare data and threaten patients' lives. In a similar context, the 2022 ransomware cyber-attack on Versailles André-Mignot Hospital compromised the healthcare system and disclosed tremendous amounts of patient information. Towards this direction, most researchers have solely developed either machine learning or Deep learning algorithms to identify network traffic anomalies. Motivated by the above challenges, an effort has been made in this paper to design a Recursive Feature Elimination (RFE) integrated with machine learning paradigms and a Ridge regression merged into deep learning models for implementing accurate anomaly intrusion detection based on the real-time dataset WUSTL-EHMS. Among the paradigms used, the proposed approach confirms that the RFE-based Decision Tree (DT) outperforms state-of-the-art techniques with a training accuracy of 99 % and a testing accuracy of 97.85 % while maintaining a reduction of FAR to 0.03. In a nutshell, it has been proven that the suggested framework can be deployed to build anomaly intrusion detection, reinforcing IoMT against widespread cyber-attacks and safeguarding the integrity of advanced healthcare systems.

1. Introduction

The revolutionary progress of Industry 4.0 was invented in 2011 by the German Government to underscore its focus on a technologically advanced approach [1]. Industry 4.0 intends to integrate advanced software, various machines, and sensor components via wired/wireless networks to control and sustain the manufacturing process. The smart city is the cornerstone category of Industry 4.0. Various countries are embracing smart cities to generate a network of connected devices that amalgamates the Internet of Things (IoT). The IoT, introduced by Kevin Ashton in 1999 [2], is a network category that links various objects to the internet via data-sensing devices facilitating data exchange, smart detection, tracing, and surveillance [3]. IoT is predominantly used in smart home applications, healthcare monitoring systems, and smart devices. It facilitates the collection and analysis of detailed information, thereby improving the quality of services provided to users [4]. According to Gartner [5], the count of connected IoT equipment is anticipated to attain 27 billion by 2025, nearly twice the amount of IoT devices connected to the internet in 2021. Subsequently, the integration of IoT intelligence capabilities into the medical equipment's branch is known as the Internet of Medical Things (IoMT), it is also noted that IoMT devices cover about 30 % of the IoT appliances market [5] bringing benefits not only for doctors and patients but also becoming an economic boon. It is anticipated that the healthcare sector will conserve around USD 300 billion annually [1]. The medical and healthcare industries are experiencing an unparalleled and swift shift towards digitalization, poised to revolutionize patient care. An illustrative instance of this transformation lies in recognition of the imperative for uncomplicated and digital tools to remotely address the psychological well-being of healthcare professionals amidst the COVID-19 pandemic [6]. The current scenario heightens the allure for cyber attackers to target the healthcare sector, particularly amidst the COVID-19 pandemic. A surge in cyber-attacks has been observed targeting numerous healthcare institutions, including hospitals, as well as compromising patient and clinical data, medical firms, universities, and laboratories. Concurrently with the brisk expansion of IoMT systems and devices, malicious cyber-attacks such as malware infections, denial of service (DOS), ransomware, MIT M, sniffing, replay, relay, and data

* Corresponding author.
E-mail address: ghita.lazrek@usmba.ac.ma (G. Lazrek).

https://doi.org/10.1016/j.rineng.2024.102659
Received 7 June 2024; Received in revised form 22 July 2024; Accepted 30 July 2024
Available online 2 August 2024

modification [7] are performed to compromise healthcare devices, pose threats to the IoMT system, facilitate illegal activities [8] and put patient's lives in danger. For instance, if trespassers gain control of intelligent pacemakers, they could administer a shock to the patient, putting their life at risk. Meanwhile, there is a risk that arises when someone deliberately manipulates insulin pumps that are connected causing an excessive administration of insulin. This could potentially result in harm or even the loss of a patient's life.

Additionally due to the critical sensitivity of IoMT data [9]. accentuated that the average cost of medical information is roughly 50 times more than other types of financial data, which makes it highly valuable on the black market. With over 12 million patient record breaches (a 300 % surge since 2015), plenty of which are reported as being available for purchase on the dark web [10]. Help Net Security, in turn, has confirmed that cybercriminals successfully breached Singapore's health system, gaining unauthorized access to private information belonging to 1.5 million patients. Furthermore, they compromised outpatient prescription data for 160,000 individuals, among them Singapore's Prime Minister [3]. In Norway's Health South-East (RHF), 2.9 million subscribers were impacted due to a cyber-attack, whereas the WannaCry ransomware attack targeting England's National Health Service led to the annulment of approximately 19,000 appointments. Moreover, to recover from these cybersecurity threats, a total of 92 million was paid out [1]. Meanwhile, cybercriminals and unauthorized access have directed their focus towards healthcare organizations, as seen in the 2018 Ransomware cyberattack that inflicted a $55,000 cost on Indiana hospitals [11]. It is evident that security is of utmost importance in IoMT solutions and must be addressed as a top priority. This underscores the growing challenges for researchers and developers in establishing trusted IoT environments, given the escalating concerns about the reliability and trustworthiness of IoT-based systems [2].

Currently, some researchers have started exploring several security mechanisms including Radio Frequency Identification (RFID) [12,13], Elliptic Curve Cryptography (ECC) [14], Homomorphic encryption [15], and Physically Unclonable Function (PUF) [1], to ensure the CIA requirements (Confidentiality, Integrity, Availability) for IoMT in the data level [16]. Unfortunately, the scalability of these traditional security techniques is limited, and their high computational resource requirements make them unsuitable for the constrained resources of IoMT amenities, it's relatively straightforward to make these devices unreachable by exhausting their battery with a severe effect. However, the emergence of computing and processing capabilities, enables other scientists to leverage Machine Learning (ML) or Deep Learning (DL) for intrusion detection. While the majority neglect to merge ML and DL in the same framework and focus solely on network traffic. Thus, in this paper, an innovative approach is designed to elevate the identification of intrusions within IoMT by employing anomaly detection, without requiring prior knowledge of specific attack signatures, and solely based on the exhibited characteristics of network traffic and patient physiological data. The work consists of Feature Elimination (RFE) integrated with machine learning paradigms and a Ridge regression merged into a deep learning model. Further, the most IoMT-suitable WUSTL-EHMS dataset is used to train and test the efficacy of all ML and DL models considered for investigation. Judiciously this study aims to boost the integrity and security of healthcare data over the IoMT network. In the end, the primary contributions of this paper can be encapsulated as the following.

1. Propose ML/DL-based RFE-Ridge models as AIDS using the network traffic and biometric patient data.
2. Perform a detailed performance evaluation.
3. Obtain outweigh performances compared to baseline works.
4. The experimental results underscored the significance of RFE feature selection as it consistently delivered superior results considering accuracy, precision, recall, FAR, AUC, and MCC.

The rest of this paper adheres to the outlined structure: Section.2 provides an overview of the context and relevant prior research. Section.3 elucidates the proposed approach and the dataset used. The experimental setup and standard evaluation metrics are investigated in Section.4. Section.5 includes the results and discussion, and finally, Section.6 wraps up the paper by exploring future directions.

2. Related work

Given the sensitive and confidential nature of patient data, privacy, and safety are underlying tasks in IoMT. Numerous authors highlighted network intrusion detection (NIDS) issues in response to their increased importance in the contemporary age of sophisticated cyber threats. In this section, the works pertaining to the detection of cyber-attacks using machine and deep learning methods in IoMT are discussed along with the constraints observed in current state-of-the-art approaches.

Reference [17] suggests a deep learning approach called the deep belief network (DBN) algorithm model for constructing a multi-classification intrusion detection over the IoMT network based on the CICIDS 2017 dataset. The proposed approach outperforms other Support Vector Machine (SVMIDS), Spiking Neural Networks (SNNIDS), and Federated Neural Networks (FNNIDS) in handling multiple attack types (DOS/DDOS, web attack, port scan, infiltration, BOT) [18]. proposed a hybrid Principal Component Analysis (PCA) Grey Wolf Optimization (GWO) based Deep Neural Network (DNN) classifier model to handle large data and faster detect cyber-attacks using the NSL-KDD dataset. An ensemble learning approach combines Decision Tree (DT), Naive Bayes (NB), Random Forest (RF) as individual learners, and XGBoost as a meta-classifier, which was deployed as Software as a Service (SaaS) on fog nodes and Infrastructure as a Service (IaaS) on the cloud side based on TON-IoT dataset, to mitigate various security threats in IoMT network [19]. The authors [20] have developed a Genetic Algorithm (GA) based RF for identifying the most optimal features that significantly contribute to intrusion detection while disregarding any irrelevant ones to minimize the high amount of inspection time, in combination with RF to perform binary, and multi-classification detection using NSL-KDD dataset. The experimentation results confirm the importance of GA-RF which gave better results compared to GA-NB and GA-Logistic Regression (LR).

An anomaly intrusion detection system was implemented by Ref. [21] using the K-Nearest Neighbors (KNN), RF, DT, NB, LR, and SVM algorithms. The purpose of their study is to fortify the IoMT environment against malicious traffic in the collected data based on TON-IoT Telemetry and IoT/IIoT datasets. After extensive experiences, the authors affirm that KNN, RF, and DT are viable approaches for intrusion detection in IoMT [3]. suggested an investigation of Deep Recurrent Neural Network (DRNN) and supervised algorithms (RF, DT, KNN, and ridge classifier) on the NSL-KDD dataset to develop an efficient IDS based on Particle Swarm Optimization (PSO). Furthermore, this paper corroborates that PSO-RF outperforms the other models. An SDN-enabled malware detection framework [22] incorporating a hybrid deep learning (DL) architecture that merges Convolution Neural Network (CNN) and Long-Short-Term Memory (LSTM) was developed to identify malware against IoMT devices based on the IoT malware dataset. An XSRU-IoMT based bidirectional (BID)-Simple Recurrent Unit (SRU) was developed to identify threats and examine the interpretability aspect of the predictive model, which ensures the trustworthiness of the healthcare system and avoids the black-box nature of ML/DL models [9].

The article [23] investigates ML models (DT, NB, KNN, Artificial Neural Network (ANN), and SVM) to identify IoMT assaults using the BoT-IoT dataset and confirm that the DT is the best paradigm. Moreover, a tree classifier using random forest-data augmentation was designed by Ref. [24], based on the WUSTL-EHMS dataset to secure IoMT architecture, nonetheless, the data augmentation produces an overfitting problem. To uncover insider threats in the IoMT cloud server [25], generated an anomaly hybrid intrusion detection based on feature selection (GA, PSO, and Differential Evolution (DE)) in conjunction with KNN and DT. Using the NSL-KDD dataset, a performance evaluation was discussed and validated the importance of GA-DT in building a detection system [5]. evaluates the performance of PSO based on numerous ML/DL classifiers including LR, RF, AdaBoost, SVM, DT, KNN, DNN, CNN, and LSTM. The evaluation was carried out on the WUSTL-EHMS dataset. The findings of the study prove the superiority of PSO-DNN across all models. Thus, a meticulous analysis indicates that this method is intricate and requires further performance improvement.

[10] has presented a supervised machine learning algorithms (SVM, DT, RF) to detect internal threats in Electronic Healthcare Record (EHR), with a recommendation to adopt an SVM [26]. leverages the power of the DL technique to secure Bluetooth in the IoMT system, hence numerous experimentations have been done using supervised, unsupervised ML models, and DNN using the BR/EDR and BLE datasets. The results show that the DNN is the best model to be applied on the second layer of the IoMT edge node, which is resource-restricted and has limited storage and processing power. Moreover, the work elaborated by Ref. [27] proposed a novel framework, DeepCAD, which involves training a standalone DNN model combined with anomaly detection rules for classification and anomaly detection in smart healthcare system (SHS).

Furthermore [28], developed an Empirical Intelligent Agent (EIA) based on a unique Swarm-Neural Network (Swarm-NN) method to detect attackers in the edge-centric IoMT framework. The presented model is assessed on the ToN-IoT dataset and proves high accuracy over other classification models [29]. introduced a swarm-neural network model for detecting intruders within a data-centric IoMT-based system. Using the real-time NF-ToN-IoT dataset, the experimental findings demonstrate the effectiveness of the proposed work over other classification models [30]. deployed machine-learning and deep-learning methods to predict unforeseen threats, and the Harris Hawk Optimization (HHO) algorithm was deployed to select optimal features based on the NSL-KDD dataset. Consequently, RNN is more effective and may achieves better outcomes compared to other exploited models [31]. aims to detect botnet attacks in IoMT using feature engineering PCA and linear discriminant analysis (LDA)-based ML and DL models including naive Bayes, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, Single-Layer Perceptron, Convolution Neural Network, and Multi-Layer Perceptron. For the experimentation phase, the N-BaIoT dataset is utilized to train and test the aforementioned algorithms. Among them, RF performed better when used with PCA. However, the PCA goal is to preserve variation as possible in the data, yet it doesn't ensure that the principal components retained are the most informative for a particular predictive task, potentially leading to information loss [32]. have created a metaheuristic algorithm like Lion optimization, Whale optimization, Spider-Monkey optimization, and Salp Swarm optimization to identify online attacks. The LSSOA framework offers a robust tool for detecting and preventing cyber-attacks in the IoMT environment, and the proposed approach proved high accuracy compared to other baselines.

Alternatively [33], proposed the use of an Enhanced Random Forest Classifier for accomplishing the Best Execution Time (ERF-ABE) for detecting DoS and delay attacks in IoMT networks. This classifier merges the advantages of random forests with optimization methods to augment performance. As per the results, implementing ERF-ABE has led to a reduction in execution time. In addition, SmartHealth is proposed by Ref. [34] to protect IoMT devices in smart healthcare system based on ML framework including KNN, RF, DT, and ANN to differentiate wicked activities in IoMT using a created dataset. From the experimental outcomes, it has been noted that SmartHealth identifies malicious activities with an accuracy of 92 % [35]. suggested a SafetyMed system integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to protect IoMT devices from malicious image data and sequential network traffic based on the CIC-IDS2017 and Maling datasets. The proposed model shows a potential accuracy level with average precision and recall [36]. proposed a Machine Learning (ML) oriented Intrusion Detection System (IDS) to detect cybersecurity breaches aimed at IoMT systems. Several ML models, such as Multinomial Naive Bayes, Logistic Regression, Logistic Regression with Stochastic Gradient Descent, Linear Support Vector Classification, Decision Tree, Ensemble Voting Classifier, Bagging, Random Forest, Adaptive Boosting, Gradient Boosting and Extreme Gradient Boosting, were used and evaluated on the ToN-IoT dataset, whereupon Adaptive Boosting proved superior performances compared to other works.

To identify cyber-attacks from IoMT networks, an intrusion detection system (IDS) for IoMTs utilizing Logistic Redundancy Coefficient Gradual Upweighting MIFS (LRGU-MIFS) feature selection followed by SVM, LR, RF, DT, and LSTM detection models is proposed by Ref. [37]. This approach exhibited high performance, ensuring its effectiveness in preventing IoMT attacks [38]. presents a secure system tailored for IoMT devices to combat DDoS cyberattacks targeting patient medical data, employing the average convolution layer (CNN-ACL), which exhibits outstanding performance relative to other machine learning methods based on the KDDCUP99 and CICIDS2017 datasets. Meanwhile [39], developed an RFE with multilayer perceptron (MLP) to detect intrusions based on the ECU-IoHT dataset, TON-IoT dataset, WUSTL-EHMS, and the ICU datasets [40]. proposed ML and DL models including the linear support vector machine (LinSVM), the convolutional support vector machine (ConvSVM), and the categorical embedding (CatEmb) to detect IoMT intrusions. The proposed DL models prove superior performance, achieving accuracies of 99.78 %, and 99.98 %, for LinSVM and ConvSVM respectively. The CatEmb model, highlighted as pioneering the use of a deep learning-based embedding approach, realizes an accuracy of 99.84 %. A novel meta-intrusion detection system that utilizes meta-learning techniques to improve detection capabilities for both known and zero-day intrusions was developed by Ref. [41]. Rigorous experimentation on (WUSTL-EHMS-2020, IoTID20, and WUSTL-IIOT-2021) demonstrated 99.47 %, 99.98 %, and 99.99 % for anomaly detection, and 99.57 %, 99.93 %, and 99.99 % for signature-based detection with low misclassification rates, affirming the meta-IDS reliability.

Based on a comprehensive review of the literature, it is evident that most studies in the field concentrate on developing IDS classification models for IoMT without adequately considering the amalgamation of ML and Deep Learning (DL) techniques as AIDS within the same framework. Furthermore, these security mechanisms often rely on outdated datasets to evaluate their models, which fail to capture modern IoMT attacks and do not align with the evolving IoMT architecture. Table 1 contrasts the ML and DL-related work in the context of detecting attacks in IoMT.

3. Proposed approach and dataset description

3.1. IoMT security framework deployment and integration

In Access Control and Authentication IoMT Security Framework, the RFE/Ridge can identify key features differentiating legitimate users from intruders, by analyzing various features such as login times, device usage patterns, transaction histories, and access locations. However, the suggested ML/DL models dynamically adjust authentication mechanisms based on real-time analysis of behavioral data, they continuously monitor ongoing user activities and compare them against established baselines. When deviations or anomalies are detected, they trigger alerts to notify security personnel or automated systems. These alerts prompt immediate action to verify the user's identity or restrict access until further verification is completed. Moreover, The Security Information and Event Management (SIEM) system is a central hub that consolidates data from various sources. It correlates events across different layers of network traffic and security detection systems, including intrusion detection systems (IDS), anti-malware software, and other security tools

Table 1
State-of-the-art attack detection in IoMT using ML and DL mechanisms.

Paper	Dataset	Method	Limitations
[17]	CICIDS 2017	DBN algorithm model	Lack investigation of ML, only network traffic.
[18]	NSL-KDD	PCA-GWO based DNN	Only usage of DL model, PCA inherently involves a loss of information, the computational cost of performing PCA is high.
[19]	TON-IOT	Ensemble learning	Only network traffic, high FAR.
[20]	NSL-KDD	GA based RF	Lack investigation of DL, GA is computationally expensive, without the adaptability of RF to changed nature of features.
[21]	TON-IoT, IoT/IIoT	Investigation of ML models (KNN, RF, DT, LR, NB, SVM), DT, RF, KNN are the suitable models	DL paradigms are not considered with ML models for intrusion detection.
[3]	NSL KDD dataset	Ridge classifier, DT, RF, KNN, RNN based PSO, RF is the best model	Dataset includes only network traffic, PSO converges prematurely, PSO is computationally expensive, without the adaptability of ML models and RNN to changed nature of features.
[22]	IoT malware	Hybrid DL-SDN malware detection	Dataset contains only network traffic.
[9]	TON-IOT	XSRU-IoMT based on BID-SRU	Dataset contains only network traffic, not considering ML techniques.
[23]	BoT-IoT	Investigation of ML approach (DT, KNN, ANN, NB, SVM), DT is the best model	Lack consideration of DL Algorithms for intrusion Detection, only network traffic was considered.
[24]	WUSTL-EHMS	Data augmentation-RF	Lack consideration of DL Algorithms for intrusion Detection, overfitting issues.
[25]	NSL-KDD	GA, PSO, DE based KNN and DT, GA-DT is the best model	Lack consideration of DL Algorithms, PSO converges prematurely, PSO/GA are computationally expensive, without the adaptability of ML models to changed nature of features.
[5]	WUSTL-EHMS	PSO based LR, RF, AdaBoost, SVM, DT, KNN, DNN, CNN, LSTM, PSO-DNN is the best model	PSO converges prematurely, PSO is computationally expensive, without the adaptability of DNN model to changed nature of features, performance improvement.
[10]	EHR-UK	SVM, DT, RF, SVM is the best model	Lack consideration of DL, not publicly available dataset
[26]	BR/EDR and BLE	DNN model	IoMT edge node has limited power and storage.
[27]	Pima Indians Diabetes Parkinson	DeepCAD	The performances still need to be improved.
[28]	TON-IOT	Empirical Intelligent Agent (EIA) based Swarm-Neural Network Swarm-NN	Only usage of DL model, EIA is Resource Intensiveness.
[29]	NF-ToN-IoT	Swarm-neural network model	Only DL model, performances still need to be improved.
[30]	NSL KDD dataset	HHO-RNN model.	HHO may stick in local optima, Without the adaptability of RNN to changed nature of features, HHO is computationally expensive.
[31]	N-BaIoT	PCA and linear discriminant analysis (LDA) based ML and DL models	PCA inherently involves a loss of information, the computational cost of performing PCA is high.
[32]	Created dataset	LSSOA	LSSOA require significant computational resources. Invalid Positive/Negative Rate (IPR/INR) still need to be reduced.
[33]	Generated dataset	Enhanced Random Forest (ERF)	Lack consideration of DL Algorithms for intrusion Detection, PCA inherently involves a loss of information.
[34]	Generated dataset.	KNN, RF, DT, and ANN	Accuracy still needs to be enhanced, only ML models.
[35]	CIC-IDS2017, Malimg	CNN-LSTM	The model is complex.
[36]	ToN-IoT	Several ML models	Only ML models.
[37]	WUSTL-EHMS-2020	LRGU-MIFS.	The accuracy still needs to be improved. The authors' focus on accuracy metric may limit understanding of the model's capabilities.
[38]	KDDCUP99, CICIDS2017	CNN-ACL.	The recall, precision, and accuracy metrics still need to be improved
[39]	ECU-IoHT, TON-IoT WUSTL-EHMS, ICU	RFE-MLP	Lack usage of DL models. without the adaptability of MLP to changed nature of features.
[40]	WUSTL-EHMS-2020	LinSVM, ConvSVM and CatEmb	CatEmb is prone to overfitting.
[41]	WUSTL-EHMS-2020, IoTID20, and WUSTL-IIOT-2021	Meta-IDS	Meta-learning can be computationally expensive.

[42]. Integrating a robust IoMT intrusion detection framework into an SIEM system is critical for enhancing security in healthcare environments. By establishing seamless data feeds or APIs, the framework can transmit alerts and security events to the SIEM, ensuring centralized monitoring and correlation with other network activities. This integration enables healthcare organizations to gain comprehensive visibility into IoMT device behavior and potential security threats in real-time. Through normalized event data and contextual enrichment within the SIEM platform, analysts can prioritize alerts, escalate incidents promptly, and initiate appropriate response actions. This unified approach not only improves incident detection and response times but also supports compliance with healthcare regulations by facilitating detailed reporting and audit capabilities. Effective integration and ongoing testing ensure that this proposed IoMT security framework operates efficiently within the broader security infrastructure, safeguarding patient data and enhancing overall cybersecurity resilience.

3.2. Proposed method

To optimize the detection process for maximum efficiency. Fig. 1

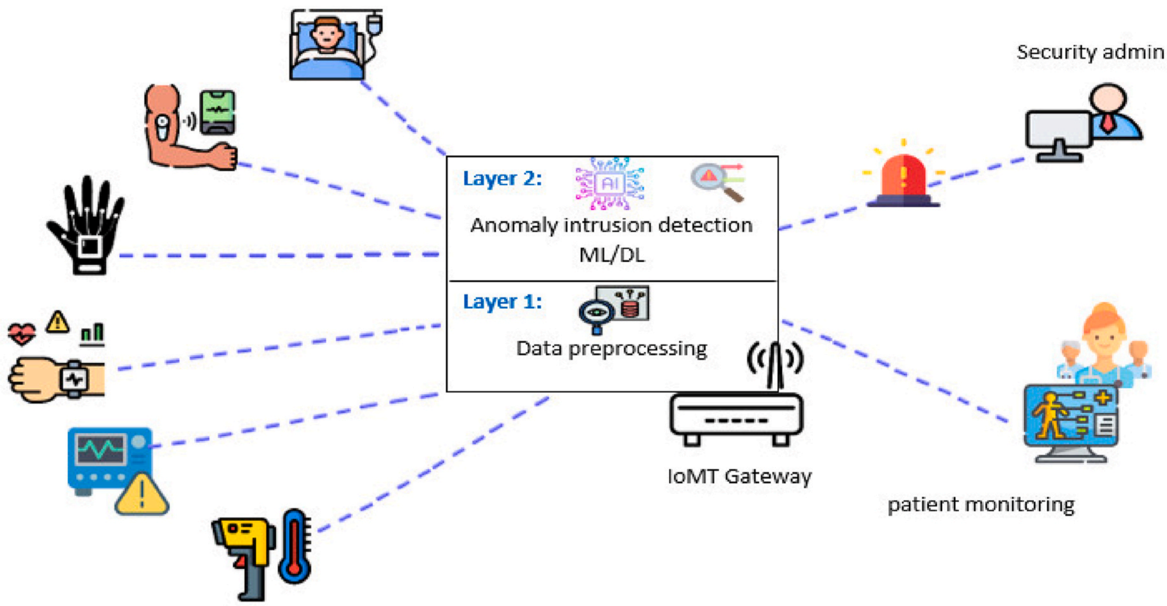

Fig. 1. Schema of the proposed ML/DL intrusion detection framework [43].

introduces a novel method for identifying threats on the Internet of Medical Things (IoMT) by leveraging network traffic data and patients' biometric information. The proposed IoMT architecture consists of IoMT devices, which are responsible for data collection and log generation. These logs hold critical information about the patient's biometric data as well as connected medical amenities. The next step involves the transmission of these logs to a centralized network gateway, which serves as the network hub for managing and monitoring the IoMT system. At the gateway, a two-layer intrusion detection system is implemented, compromising a data preprocessing layer (tier 1) and an anomaly detection layer (tier 2) to identify potential threats and assaults. Concerning attacks, the security admin plays a pivotal role in decision-making to block and discard malevolent data, while normal data is transmitted to the patient monitoring unit, which provides the essential treatments and the corresponding drugs to the patient.

In contrast to network infrastructure intrusion detection, this study encompasses IoT detection from IoMT systems to identify patient biometric aberrations. Fig. 2 shows the main components of the suggested intrusion detection method. Initially, the network traffic and patient biometrics are combined through the synchronization of timestamps for both the network flow events and the patients' biometric data, which produce the dataset. A "Data preprocessing engine" in layer 1 is then applied to clean, encode, split the dataset, normalize, and select relevant features. Subsequently, in the "Anomaly intrusion detection engine" at layer 2, hyperparameters-tuned ML/DL classifiers were used through grid search CV in the last phase to improve overall classification performance, followed by training ML/DL models on the selected features to detect any unusual patterns or deviations in both patient biometrics and network behavior. At the heart of this approach, we describe in depth the key functions associated with each layer.

3.2.1. Data preprocessing engine

The dataset is processed to extract valuable insights and presents the data in a compatible manner with ML/DL algorithms, facilitating rapid model training and testing and ultimately leading to increased classification accuracy. We furnish a comprehensive explanation of data preprocessing procedures below.

- **Removal of irrelevant features:** It involves identifying and eliminating features from a dataset [44] that can create noise during model training and that do not contribute significantly to the predictive task at hand, such as source and destination information, because in certain scenarios, the adversary may spoof these addresses, resulting in the dissemination of inaccurate or misleading information.
- **Label encoding:** Is a technique utilized to convert categorical variables into numerical values by affecting a single numerical value for each distinct category or label and keeping such numbering consistent throughout the feature to be comprehensible by machine learning [45].
- **Dataset splitting:** Is a promising step in the process of training and evaluating machine learning models. It implicates separating the dataset into various subsets, typically for training and testing portions. Dataset splitting ensures that the model is trained on one subset of data and evaluated on another to assess its performance. In this research, the envisaged dataset will be partitioned into 80 % training and 20 % testing (recommended to avoid overfitting issues).
- **Min-max scaling:** Is a data normalization technique used to scale features to a particular range, where the range of scaled values falls between 0 and 1. This technique aids in ensuring that the learning process is not biased towards features with larger magnitudes compared to features with smaller magnitudes. Mathematically, the Min-max scaling can be given as per Eq. (1)

$$z = \frac{X - min(X)}{max(X) - min(X)} \tag{1}$$

Where z represents the scale value of the initial data X and respectively $min(X)$ and $max(X)$ denote the minimum and maximum values of the feature within the dataset.

- **Feature selection:** Also termed attribute selection, it is the procedure of selecting a subset of pertinent features from a large dataset to identify the most informative and discriminating features. The goal of feature selection is to enhance model performance and prevent overfitting. In this work, an RFE wrapper method [46] and a Ridge regression embedded technique [47] are used as feature selection methods.

1. **Recursive Feature Elimination (RFE):** Is a wrapper method that aims to select the most relevant features by eliminating less important ones. It starts with all the features and removes the less meaningful ones until the desired number of features remains. In this study, we employed a process called Recursive Feature Elimination

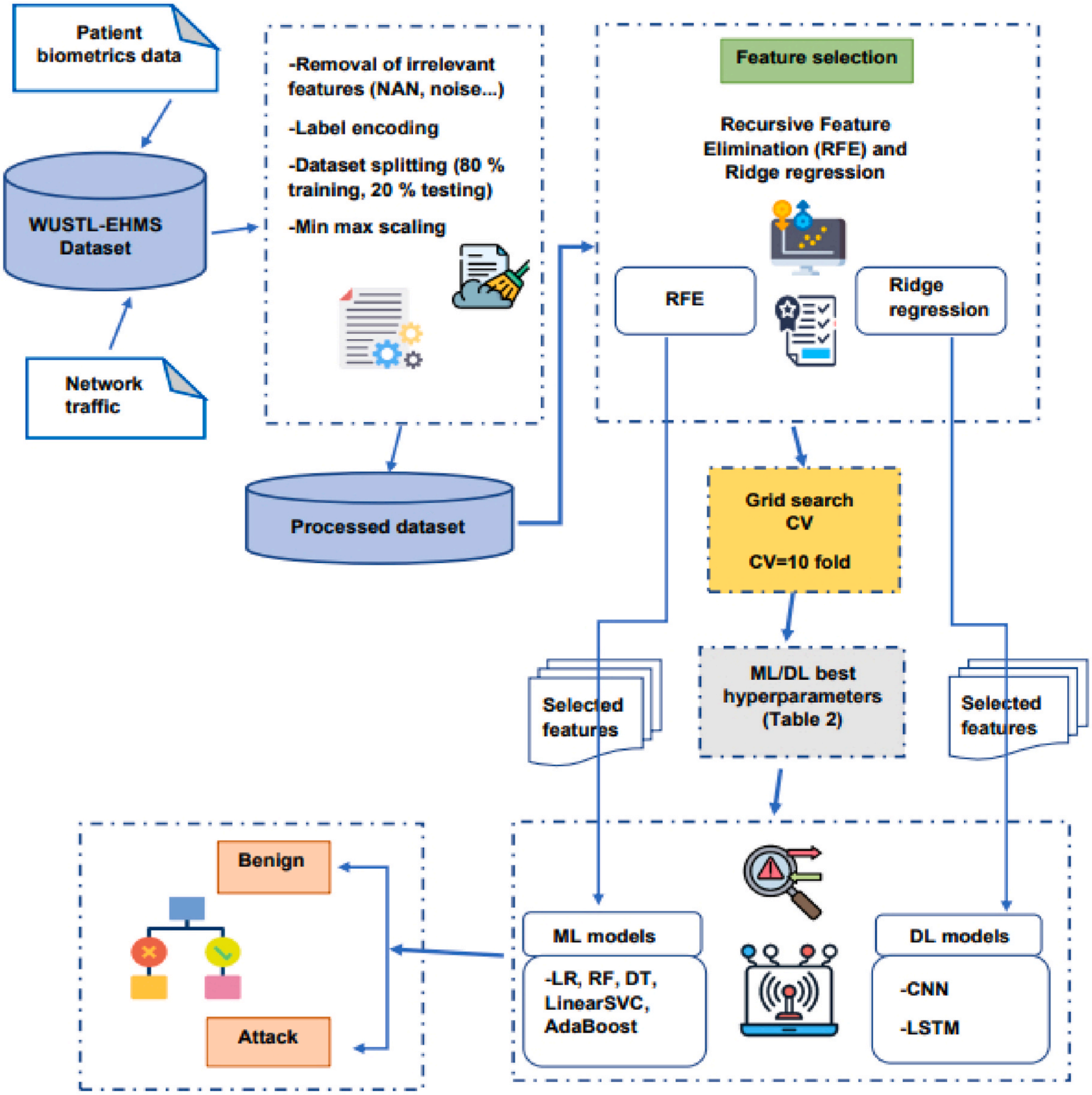

Fig. 2. Working principle of the suggested ML/DL based RFE/Ridge approach.

(RFE) with estimators such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Classifier (SVC), and AdaBoost to identify sets of eight features to personalize the feature selection. Each set was employed to train its corresponding machine-learning model. Specifically, the LR model utilized features selected by RFE using the LR estimator, while the DT model incorporated features chosen through RFE with the DT estimator. Similarly, the RF model integrated features selected by RFE using the RF estimator, and so on for the SVC and AdaBoost models.

2. **Ridge regression:** Is an embedded feature selection technique that adds an L2 penalty term to the objective function of a regression model. The regularization term in the Ridge regression helps to reduce the coefficient estimates towards zero, but unlike LASSO, it does not set any coefficients exactly to zero. This means that all features can still contribute to the model, but with smaller quantities for less important features. In this work, the features identified as most significant through the RIDGE algorithm selection were used to train the deep learning models employed in AIDS, and the regularization parameter α is set to 1.
3. **Methodology purpose:** We have used RFE integrated with various ML model estimators such as Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), LinearSVC, and AdaBoost. The outputs of RFE using RF, DT, LR, LinearSVC, and AdaBoost as estimators are used to train RF, DT, LR, LinearSVC, and AdaBoost models respectively. On the other hand, the findings of Ridge regression are incorporated into deep learning models. To the best of our knowledge, we are the first group to use a combination of RFE with ML models and Ridge with DL models in a single framework to adapt the intrusion detection models to the evolving nature of features to improve performances and minimize misclassification.

3.2.2. Anomaly intrusion detection engine

After data preprocessing in layer 1, the IoMT features selected by RFE are forwarded to layer 2 within the IoMT Gateway as inputs to train manifold machine learning models such as Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), AdaBoost, and LinearSVC. Simultaneously, the features chosen through Ridge are exclusively utilized to train deep learning models involving Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). The ML and DL algorithms utilize the selected features with the best hyperparameters obtained through Grid Search CV to differentiate between attacks and normal instances. Further, the numerous steps of this layer are explained below.

- **Grid Search Cross Validation (CV):** It is a brute-force exhaustive technique used for hyperparameter tuning to search for the optimal combination of hyperparameter values that realizes the optimal performance for a given model, furthermore, grid search works by exhaustively evaluating the model's performance for all possible combinations of hyperparameter values defined in a grid search process and selecting the best hyperparameters that produce superior performance. The parameters range utilized to initialize the grid search process with 10-fold cross-validation and the attained results are displayed in Table 2.
- **Logistic Regression (LR):** Is a statistical model widely used for binary classification tasks [48] to predict the probability of an instance affiliated with a specific class based on its features. Logistic regression intends to uncover the relationship among the input variables and the binary target variables by estimating the probabilities using a logistic function Eq. (2) known as the sigmoid function, which maps any real number into a value ranging between "0″ and "1". Furthermore, logistic regression can be (i) binary, where the dependent variable, or the output, has two possible categories, like distinguishing between benign and anomaly. (ii) For multinomial classification, the dependent variable can take on multiple categories. For instance, it might classify observations into benign, attack 1, and attack 2. (iii) Another application is ordinal logistic regression. It's a type of multinomial regression, but in this case, the categories have a specific order or ranking, such as different levels of attack severity.

$$f(x) = \frac{1}{1 + exp(-z)} \tag{2}$$

Here, z is the linear combination of features and their respective weights. To clarify, $z = w_0 + w_1x_1 + w_2x_2 + \ldots w_nx_n$. Thus, $w_0, w_1, w_2, \ldots, w_n$ are the weights and $x_1, x_2, \ldots, x_n$ are the features. In logistic regression, the model's output is determined through a threshold and decision boundary. Specifically, in a binary scenario described by Eq. (3), if the predicted probability is equal to or exceeds 0.5, the observation is designated to class A; otherwise, it is allocated to class B. This threshold-based decision rule helps categorize the predictions into the respective classes based on the likelihood estimated by the model [49].

$$Y = \begin{cases} A, if\, f(x) \geq 0.5 \\ B, otherwise \end{cases} \tag{3}$$

Table 2
Hyperparameters optimization for ML/DL algorithms.

ML/DL models	Parameters	Value range	Optimal value
Logistic Regression (LR)	Penalty: C	{0.1,1,10}	10
AdaBoost	n_estimators	{100,200,500,1000}	1000
	learning_rate	{0.001,0.01,0.1,1}	1
Random Forest (RF)	min_samples_split	{2,5,10}	5
	n_estimators	{10,50,100,200}	100
Decision Tree	criterion	{gini, entropy}	gini
	min_samples_split	{11,40,9}	11
	min_samples_leaf	{1,2,3}	1
LinearSVC	Penalty: C	{0.1,1,10}	10
	Max_iter	{1000,5000,10000}	1000
CNN	filters	{32,44}	44
	kernel_size	{3, 5}	5
	pool_size	{2,3}	2
LSTM	neurons	{50,100,150}	150
	activation	{relu, tanh}	tanh
	optimizer	{adam, rmsprop}	adam

- **AdaBoost:** Adaptive Boosting Classifier, or AdaBoost is a meta-estimator mainly used for classification, and the base learner (the machine algorithm that is boosted) combines a multiple-week classifier to create a strong classifier [50]. AdaBoost is a decision tree with only one level (stumps). This model is recognized for its ability to avoid overfitting, deal with noisy data, and accomplish high accuracy on several classification tasks.
- **Random Forest (RF):** Is a popular machine-learning model that is used for both classification and regression tasks. Thus, RF is an ensemble learning method that amalgamates numerous decision trees [51] to build a viable model. This approach enhances overall accuracy and stability when compared to a single decision tree. What sets RF apart is its ability to effectively manage randomness in both samples and features. During the prediction process for a new sample, the model aggregates predictions from each tree within the forest. The result is determined by employing an averaging or voting mechanism, consolidating the diverse insights from the individual trees.

Random Forest, an ensemble learning technique, employs multiple decision trees to enhance the reliability and accuracy of classification or regression tasks. It involves an ensemble of decision trees, where each tree is trained on distinct subsets of both the training data and features. The ultimate prediction is produced by combining the outputs of these trees. The decision rules and outcome probabilities of both individual decision trees and the ensemble are represented through a set of equations or inequalities, serving as the mathematical representation of the Random Forest algorithm [52].

With n decision trees and m features, Let F represent a Random Forest. Each decision tree i is trained using a subset of features F_i and training data D_i. For a given set of branches $B_{i,j}$ representing potential outcomes or their probabilities, each node j of tree i symbolizes a decision or a chance event, denoted by $D_{i,j}$. Each branch k emanating from node j leads to a child node l and is associated with a probability of occurrence $p_{i,j,k}$. The probabilities of outcomes in a decision tree can be depicted through either conditional probabilities or joint probabilities, explaining the likelihood of each result considering the decision rules and preceding outcomes. Let b be a constant value, and Y represents a random variable, signifying the target variable or class label of the data. Given the decision rule $D_{i,j}$ and the prior outcome $Y_{i,j-1}$ the probability of outcome k for node j of tree i can be expressed as follows Eq. (4):

$$p_{i,j,k} = P\left(Y_{i,j} = k \mid D_{i,j}, Y_{i,j-1} = b\right) \tag{4}$$

The probabilities generated by all decision trees can be aggregated to obtain the probability of the Random Forest result. For classification tasks and regression tasks, the most commonly used aggregation methods are the mean or median. Let $y_{i,}$ represent the expected value of tree i for a certain instance. The forecasted value of the Random Forest can be expressed as outlined Eq. (5):

$$y_{RF} = vote\ (y_1, y_2, \ldots y_n) \tag{5}$$

Random Forest (RF) is widely adopted for classification and regression tasks due to its simplicity, reduced susceptibility to overfitting compared to individual decision trees, and its capability to manage high-dimensional datasets containing numerous features [53] leveraging a random forest classifier for monitoring presents a spectrum of potential benefits. These encompass heightened accuracy, resilience to noise, the ability to conduct feature importance analysis, adept handling of intricate relationships, and scalability.

- **Decision tree (DT):** Is a machine learning algorithm [54] commonly employed for classification and regression problems. This model applies a flowchart-like tree structure where each internal node indicates the feature, branches denote the rules, and the leaf nodes designate the result of the algorithm. Generally, DT is fast to train, able to deal with large datasets, and can effectively capture the nonlinear relationships between features and the target variable. Decision Trees find extensive application across diverse domains such as finance, healthcare, and marketing, facilitating data-driven decision-making processes [55].

In Decision Tree (DT) models, the root node is positioned at the top, and the branches, determined by the essential characteristics of the data, extend downward. Output branches signify the output of a feature, while output child nodes represent the output of specific categories within the tree structure. Learning this classification model involves using a classification Decision Tree, which serves as an example of supervised learning. This approach depends on sample training, where the model learns from labeled examples. The classification process reaches its conclusion once the incoming data evaluated by each node have been meticulously examined [56]. This model exhibits resilience to data quality variations and proves highly proficient in processing extensive volumes of network traffic data due to its rapid data learning rate. Its applicability is enhanced by the transparency of the white-box model, which allows for the interpretation of judgment results. This makes it well-suited for detecting attacks that are particularly vulnerable to False Alarms and analyzing attack traffic data, which may provide insights into the intentions of the attackers. Furthermore, the decision tree model possesses the capability to accurately assess data impurity through the examination of entropy, denoted as $\sum_{i=1}^{C} - f_i \, log(f_i)$. Moreover, the Gini coefficient $1 - \sum_{i=1}^{C} f_i(1 - f_i)$, is employed to evaluate data impurity by considering the frequencies f_i of each data label i within the unique label set C. This function proves highly effective in noise reduction within the dataset [57].

- **LinearSVC:** Is a variant of Support Vector Machine (SVM) tailored for binary classification, in particular, to find the best hyperplane [58] that separates the two classes Eq. (6). LinearSVC is valuable in handling large datasets, and high numbers of training samples. This algorithm boasts exceptional accuracy, showcasing its effectiveness across both binary (two classes) and multi-class (more than two classes) classification tasks. Its prowess has been thoroughly validated through extensive demonstrations across diverse scenarios.

$$w * x + b = 0 \tag{6}$$

Where w is the weight vector, x is the input feature vector, and b is the bias term that moves the hyperplane away from the origin.

- **Convolutional Neural Network (CNN):** It is a category of deep learning algorithms [59] that is utilized specifically for image recognition and computer vision tasks, as well as sequential data involving time series data or text data. CNN has an input layer, an output layer, and numerous hidden layers, allowing it to learn complected objects and patterns. The Conv2D layer is widely used for computer vision and other tasks, including two-dimensional data processing, in contrast to the Conv1D layer, which operates in a one-dimensional layer to process sequences or time-series data [60]. The operation requires applying a kernel to the input sequence. The filter is moved along the sequence, and at each position, the elements of the filter are multiplied with corresponding elements of the sequence. These results are then summed up to generate a value in a sequence, which is referred to as a feature map. This paper uses one convolutional layer, one max-pooling layer, one flatten layer, and two dense layers, as shown in Fig. 3. The convolution operation in the convolution layer helps the feature extraction, and the max pooling layer helps to capture the maximum activated features. The result obtained from the max pooling layer acts as the input for the flattening layer, which is responsible for adjusting the data shape to match the requirements of the subsequent fully connected layer. Toward the conclusion of the network, we employ two fully connected layers. The latest dense layer (output) number of neurons matches the number of output parameters, which in our scenario is 1, due to the desired binary classification outcome.
- **Long Short-Term Memory (LSTM):** Is a category of Recurrent Neural Network, used for sequence modeling tasks. LSTM possesses the ability to learn long-term dependencies [61] and overcome vanishing gradient problems. Moreover, it comprises a structure consisting of four interacting layers, as illustrated in Fig. 4. The cell structure can be simplified into three parts: a forget gate (f_t), an input gate (i_t), and an output gate (o_t). The input gate and output gate are responsible for handling the input and output of data at a given time. The forget gate compares the data input against the previous data state to decide which information to discard [62]. The relationship between the gates in an LSTM cell can be described mathematically as follows in Eqs. (7), (8), (9), (10), (11) and (12)

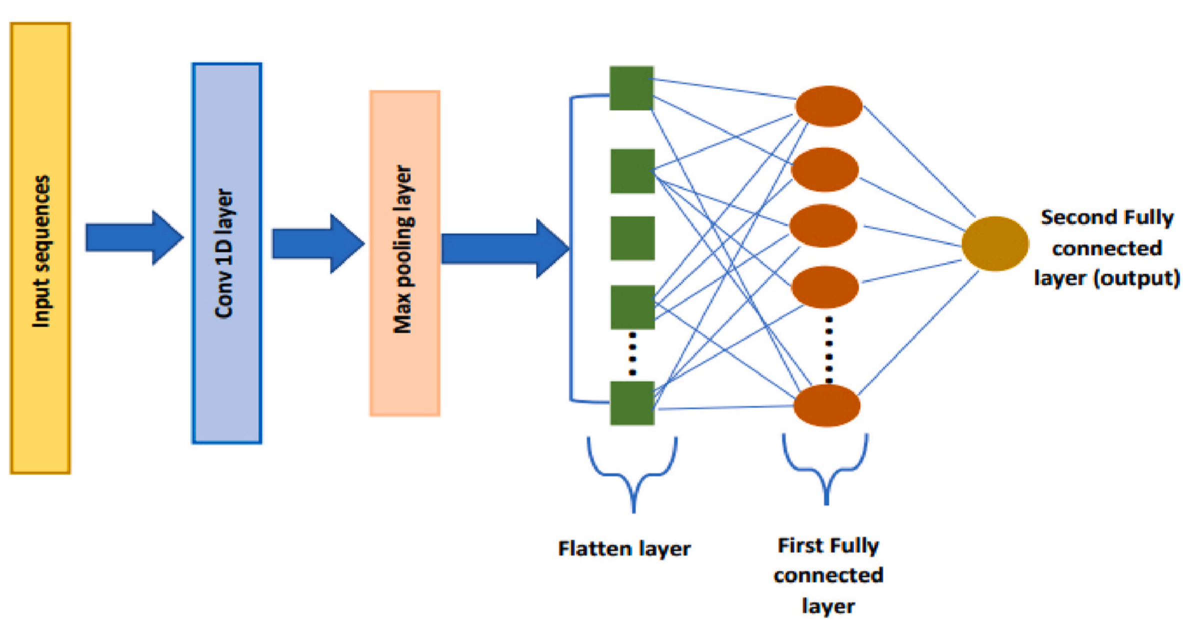

Fig. 3. CNN model architecture.

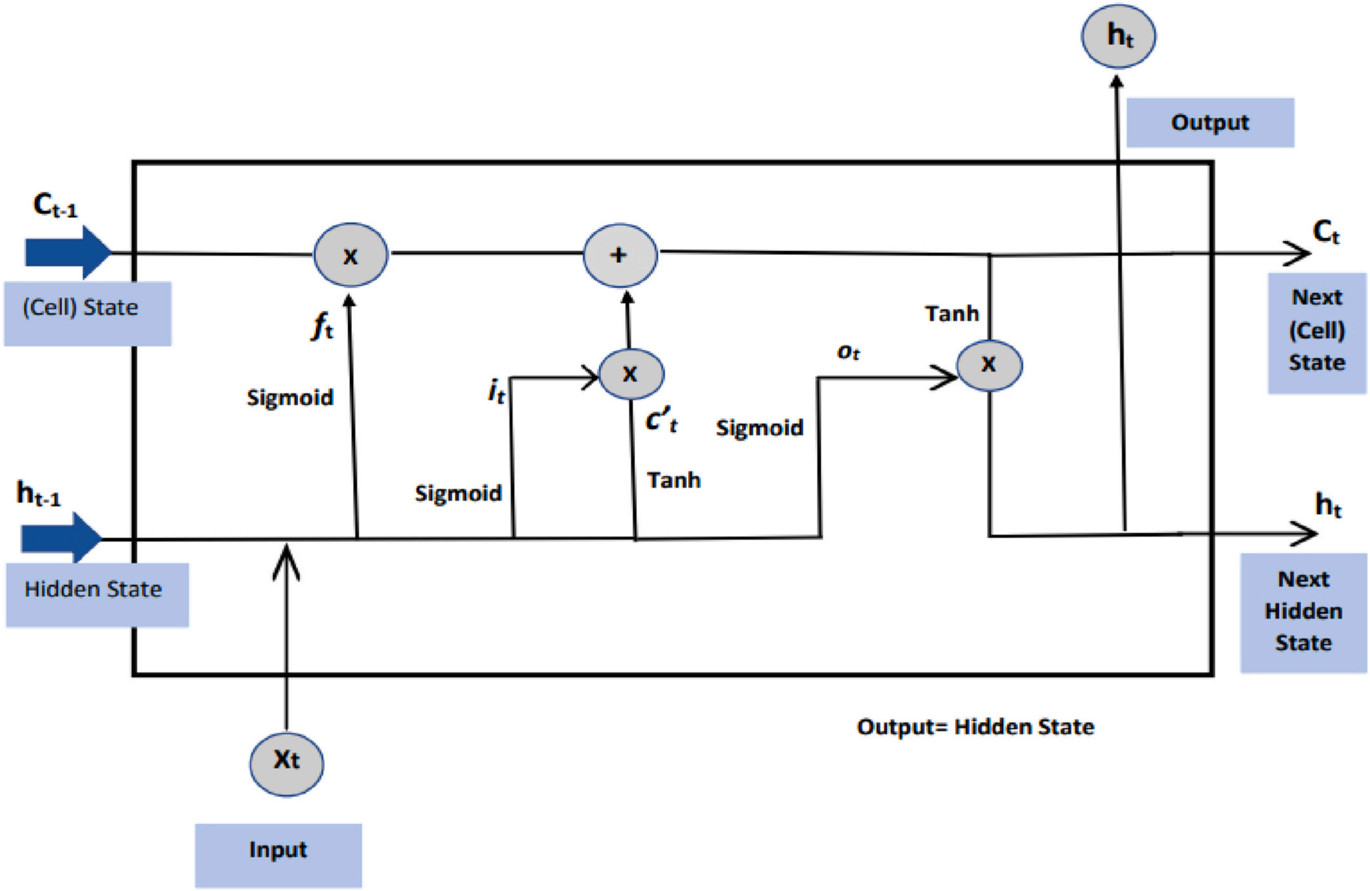

Fig. 4. LSTM cell structure.

[63], where (w_i), (w_f), (w_o) and (w_c) are the weight matrices, (b_i), (b_f), (b_o), and (b_c) are the biases. As seen in Fig. 5 our model contains three LSTM layers, three dropouts, and two fully connected layers. The accuracy value and binary cross-entropy loss function were used as the fundamental metrics. The dropouts are applied to prevent overfitting during the training process. In our model, the dropouts have a value of 0.2, which means that during each training 20 % of the neurons will be randomly deactivated.

$$i_t = \sigma(w_i \bullet [h_{t-1}, x_t] + b_i) \tag{7}$$

$$f_t = \sigma\left(w_f \bullet [h_{t-1}, x_t] + b_f\right) \tag{8}$$

$$o_t = \sigma(w_o \bullet [h_{t-1}, x_t] + b_o) \tag{9}$$

$$c_t' = tanh(w_c \bullet [h_{t-1}, x_t] + b_c) \tag{10}$$

$$c_t = f_t * c_{t-1} + i_t * \ c_t' \tag{11}$$

$$h_t = o_t * tanh(c_t) \tag{12}$$

3.3. Dataset description

The WUSTL-EHMS dataset-2020 [64] was formed using a real-time Enhanced Healthcare Monitoring System testbed. The testbed comprises four main elements: medical sensors, a gateway, a network, and a control and visualization module. The network flow metrics and patient biometrics are combined using the testbed, and the biometrics data

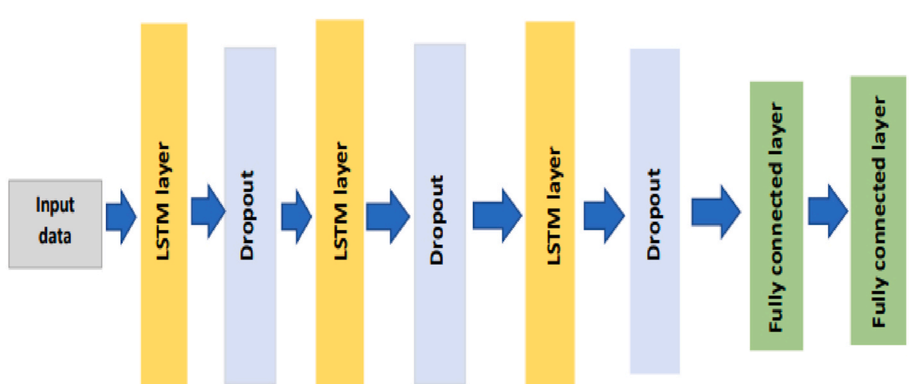

Fig. 5. LSTM model.

includes parameters such as temperature, peripheral oxygen, and saturation, which are collected using different health monitoring sensor boards. The board is connected to a Windows-based computer through a USB port. The system was created utilizing C++ to capture the sensed data. The data is then transmitted from sensors affixed to the patient's body to the gateway, which forwards it to the server for visualization. A wrongdoer may intercept the data, and the IDS is responsible for capturing network traffic and patient biometrics data, as well as detecting abnormalities. Finally, the Audit Record Generation and Utilization System (ARGUS) tool collects and stores network flow and patient data as a CSV file. The WUSTL-EHMS dataset encompasses Man in the Middle attack: spoofing and data injection, to sniff and alter packets on the fly, which respectively disrupt the confidentiality and integrity of patient data. This dataset includes 44 features, with 35 of them specifically representing network traffic metrics, eight representing patient biometric features, and one additional feature designated for the label. The attack traffic is labeled as "1″, whereas the normal traffic is indicated by "0". Table 3 displays the quantitative information on the WUSTL-EHMS dataset.

4. Experimental setup

This section describes the software and hardware configuration used, additionally, it provides the evaluation metrics and comparison methods to accurately identify attacks by considering binary classification. The simulation environment, the performance metrics used, and the proposed comparison review are mentioned below.

4.1. System setup

The experiments were performed on a Desktop using system specification Intel(R) Core (TM) i5-6200U CPU @ 2.30 GHz 2.40 GHz and 12,0 Go RAM. The approach is being trained and tested using ANACONDA NAVIGATOR 2.3.2- Jupyter Notebook [66]. All selected classifiers were built in Python using the Scikit-learn toolkit, NumPy, Matplotlib, KerasClassifier [67], GridSearchCV [68], train_test_split function, MinMax scaler [69], label encoder.

4.2. Evaluation metrics

The performance of the seven suggested models is assessed utilizing statistical parameters [70] to correctly identify intrusions by considering binary classification. We have considered four frequently utilized metrics to evaluate the performance of IoMT attack detection: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are used to outline the following metrics.

- **Accuracy:** Is presented as the ratio of correctly predicted traffic categorization (comprising both normal and attack) to the overall predictions made on the test data for network traffic using Eq. (13).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

- **Precision:** This metric measures how accurately a model can correctly identify positive (Tp) out of all the predicted positives and is defined by Eq. (14). A high precision value indicates that when the model predicts an outcome, it is more likely to be accurate. Conversely, a low precision value implies that the model could be generating false positives.

Table 3
WUSTL-EHMS information [65].

measurement	value
Dataset size	4.4 MB
Normal instances	14,272 (87.5 %)
attack instances	2046 (12.5 %)
Total samples	16,318

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

- **Recall/Detection Rate:** It is described as the proportion of the correct network attack traffic classification to the total sum of the correctly classified attack network data and misclassified network attack flow in the dataset and is given by Eq. (15).

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

- **False Acceptance Rate/False Positive Rate (FAR/FPR):** It is described as the proportion of incorrectly classified normal (non-attack) network traffic (False Positives) to the total actual normal network traffic samples in the dataset. FPR tells us how often the model produces false alarms by incorrectly classifying normal data points as malicious and is defined by Eq. (16).

$$FAR = \frac{FP}{TN + FP} \tag{16}$$

- **ROC-AUC (Area under the Receiver Operator Characteristic Curve):** This metric measures the correlation between predictions and actual labels. A higher AUC-PR score indicates the model's superior capability to differentiate positive and negative class points effectively. The AUC values are typically on a scale from 0 to 1, where:
  - An AUC of 0.5 signifies that the model performs no better than random guessing.
  - An AUC of 1 suggests that the classifier distinguishes between positive and negative instances.
- **MCC (Matthew's correlation coefficient):** Is a metric utilized to measure the performance of binary classification models in machine learning using Eq. (17). Moreover, −1 and +1 respectively indicate a perfect misclassification and classification, while MCC = 0 corresponds to the expected result for a random or "coin-tossing" classifier.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{17}$$

- **Memory:** Is a measure of the model's memory consumption in Bytes (B).

4.3. Comparison methods

The suggested approach's performance will undergo comparison against six other mechanisms that utilize WUSTL_EHMS, KDDCUP99, and CICIDS2017 datasets, namely, the approach presented by Ref. [24], the one exploited by Ref. [5], the other proposed by Ref. [39]. The LRGU-MIFS system deployed by Ref. [37], the CNN-ACL system designed by Ref. [38], and the SmartHealth framework deployed by Ref. [34]. We will now briefly explain each algorithm.

1. **Reference** [24]**:** In this paper, the authors develop a lightweight model designed for intrusion detection. It consists of Random Forest, robust scaling, and data augmentation by applying Synthetic Minority Over-sampling Technique (SMOTE) to effectively increase the representation of minority class. The method proposed utilized hyper-parameter tuning to attain the best estimator that distinguishes benign and attack samples, leading to superior performance compared to other models, notably logistic regression, decision tree, and extra tree classifier.

2. **Reference** [5]**:** The proposed framework involves an extensive examination of employing a variety of Machine Learning (ML) and Deep Learning (DL) techniques based on Particle Swarm Optimization (PSO) for network intrusion detection in IoMT and confirms that PSO-DNN achieves the best performance. The suggested method uses a PSO for feature selection, which obtains 8 optimal features from the combined set of 43 features encompassing network and biometric data during the feature selection procedure.
3. **Reference** [39]**:** In this article, the authors proposed a model for detecting cyber-attacks and anomalies in IoMT networks, employing recursive feature elimination (RFE) and a multilayer perceptron (MLP) to differentiate between normal and malicious data. The RFE method was utilized to select optimal features, employing logistic regression (LR) and extreme gradient boosting regression (XGBRegressor). The parameters of the MLP were tuned using hyperparameter optimization and 10-fold cross-validation. A series of experiments were carried out to demonstrate the efficacy of the envisaged approach in combating cyber-attacks in healthcare applications.
4. **Reference** [37]**:** The authors in this work developed an improved intrusion detection system for the IoMT network including Logistic Redundancy Coefficient Gradual Upweighting (LRGU) integrated into mutual information feature selection (MIFS). The LRGU-MIFS selects the pertinent and unique features using the LRGU redundancy coefficient. After selecting the features, they are used to train the machine learning classifier, thereby mitigating the overfitting issue potentially triggered by redundant features.
5. **Reference** [38]**:** In this work, the authors proposed an average convolution layer (CNN-ACL), a unique CNN architecture designed to learn the content features of anomalous activities and subsequently identify individual anomalies. This powerful system offers a highly detecting DDoS in the IoMT environment.
6. **Reference** [34]**:** The proposed SmartHealth system is a unique ML-based security framework designed to detect threats against smart healthcare systems (SHS). In this work, Artificial Neural Network (ANN), Decision Tree, Random Forest, and K-Nearest Neighbors algorithms are employed because they are all simple to apply in anomaly detection.

5. Results and discussion

In this section, we will comprehensively showcase the effectiveness of the proposed methodology, while also evaluating the suggested method in terms of its memory consumption in case of feature selection and anomaly detection task (grid search + model training, and prediction of ML/DL models), while considering multiple metrics to establish its superiority over existing methodologies, and its appropriateness for anomaly detection within IoMT system. This research methodology is conducted on the WUSTL-EHMS dataset, affording superior results compared to existing techniques. The experiments were fulfilled to analyze the performance metrics of the introduced model on the chosen features. The selected features performed by RFE and Ridge for any models are depicted in Table 4.

Table 4
List of RFE and Ridge selected features per models.

Estimators' models	Feature selection Method	Selected features
LR	RFE	SrcLoad, DIntPkt, DstJitter, Dur, Load, Loss, pLoss, Pulse_Rate
AdaBoost	RFE	Sport, SrcLoad,DIntPkt, SrcJitter,DstJitter, Packet_num,Pulse_Rate, Resp_Rate
RF	RFE	Sport, SrcLoad,DstLoad, DIntPkt,DstJitter, Dur,Rate, Packet_num
DT	RFE	Sport, DIntPkt,SrcJitter, DstJitter, Packet_num,Temp, Pulse_Rate,Resp_Rate
LinearSVC	RFE	DIntPkt, SIntPktAct,DstJitter,Dur,TotPkts, Loss, pLoss,pDstLoss
–	Ridge	SrcBytes, DstBytes,SrcLoad, DstLoad, DIntPkt, SrcJitter,DstJitter, dMaxPktSz,sMinPktSz, Dur,TotPkts, Load,Loss, pLoss,pSrcLoss, pDstLoss, Rate,SpO2,Pulse_Rate

5.1. Performance analysis of RFE-based ML models

The real-time WUSTL-EHMS dataset is used to evaluate the proposed approach performance using standard metrics, training accuracy (Train. acc), testing accuracy (Test.acc), precision, recall, AUC, MCC, and FAR. An RFE feature selection technique was applied to eliminate features with the lowest ranks one after the others for each model in a customized manner. Table 5 shows the detection performance for all RFE-ML models. According to the results, a crucial analysis illustrates that the five RFE-ML models achieve a high training accuracy between 92 % and 99 %, In contrast, the RFE-LR with the best penalty of 10, produced the lowest value of 92.73 % for training accuracy, RFE-AdaBoost yielded a 96.92 %, and RFE-LinearSVC realizes 92.93 %, RFE-RF obtained a 96.75 %, on another hand the RFE-DT performed a 99 % training accuracy. Besides, the RFE-ML paradigms testing accuracy range from 92 % to 98 %, a 92.61 %, 95.61 %, 92.95 %, 93.10 %, and 97.85 % respectively for RFE-LR, RFE-AdaBoost, RFE-LinearSVC, RFE-RF, and RFE-DT. Also, these ml paradigms have a precision between 93 % and 99 %, RFE-LR and RFE-Adaboost produced a precision of 93.96 %, 96.58 %, RFE-LinearSVC tandem with RFE-RF paradigms accomplish 98.94 % and 98.97 %, then the RFE-DT gains a precision of 96.50 %. Hence the recall obtained fewer values within 44 % and 86 %. The five proposed models, RFE-LR, RFE-AdaBoost, RFE-LinearSVC, RFE-RF, and RFE-DT mutually produced 44.95 %, 68.02 %, 45.19 %, 46.39 %, and 86.29 %. The AUC metric is between 72 % and 93 %, RFE-LR and RFE-LinearSVC attain 72.26 % and 72.56 %, followed by RFE-RF, RFE-AdaBoost, RFE-DT which bring 73.16 %, 83.83 %, and 92.92 %. Moreover, the five aforementioned ML models produce an MCC range from 62 % to 84 %, the RFE-LR achieves a lower MCC of 62.06 %, on the other hand, the MCC for RFE-LinearSVC, and RFE-RF were 64.26 % and 65.09 % respectively, in comparison of 81.31 % for RFE-AdaBoost and 84.02 % for RFE-DT. Finally, the FAR is measured to assess the proportion of incorrectly classified normal instances as attacks ones, the RFE-LR produced a higher FAR of 0.06, RFE-AdaBoost and RFE-DT obtained a 0.03, then RFE-LinearSVC and RFE-RF only produced a FAR of 0.01. The reason behind these results is the deployment of RFE which strongly reduces the dataset dimensionality from 44 features to 8 selected ones, and avoids noise by eliminating unnecessary features.

The results evidenced that the RFE-DT model surpasses the other four classifiers in training accuracy, testing accuracy, precision, and recall, it also achieved the highest AUC indicating a better separation between normal and attack classes, thus, a superior MCC metric which reveals a better overall model performance in dealing with imbalanced dataset in our case (14.272 normal samples, and 2.046 attacks ones) and predicting both positive and negative instances accurately. Whereas it realizes a FAR of 0.03 compared to 0.01 of RFE-LinearSVC and RFE-RF. Proceeded by RFE-AdaBoost, RFE-RF, and RFE-LinearSVC. Hence RFE-LR is the poorest that exhibits the last training/testing accuracy, precision, recall, AUC, MCC, and a high FAR which indicates a misclassification of normal instances as attacks. In other words, the LR model produces a large number of false positives which lead to unnecessary alarms, wasted resources, and potential disruption in operation. The research highlighted that the DT-RFE is the best model that successfully classifies benign samples and attacks traffic, among the other four models, except the RFE-LR which is less incapable of detecting attack instances. Lastly, it can be visually presented in Fig. 6 that the blue plot of DT-RFE outshines the orange, green, grey, and gold plots of RFE-RF, RFE-LinearSVC, RFE-AdaBoost, and RFE-LR respectively.

Table 5
Performance results for ML models.

Models	Training.acc	Testing.acc	Precision	Recall	AUC	MCC	FAR
RFE-LR	92.73 %	92.61 %	93.96 %	44.95 %	72.26 %	62.06 %	0.06
RFE-AdaBoost	96.92 %	95.61 %	96.58 %	68.02 %	83.83 %	81.31 %	0.03
RFE-LinearSVC	92.93 %	92.95 %	98.94 %	45.19 %	72.56 %	64.26 %	0.01
RFE-RF	96.75 %	93.10 %	98.97 %	46.39 %	73.16 %	65.09 %	0.01
RFE-DT	**99 %**	**97.85 %**	**96.50 %**	**86.29 %**	**92.92 %**	**84.02 %**	**0.03**

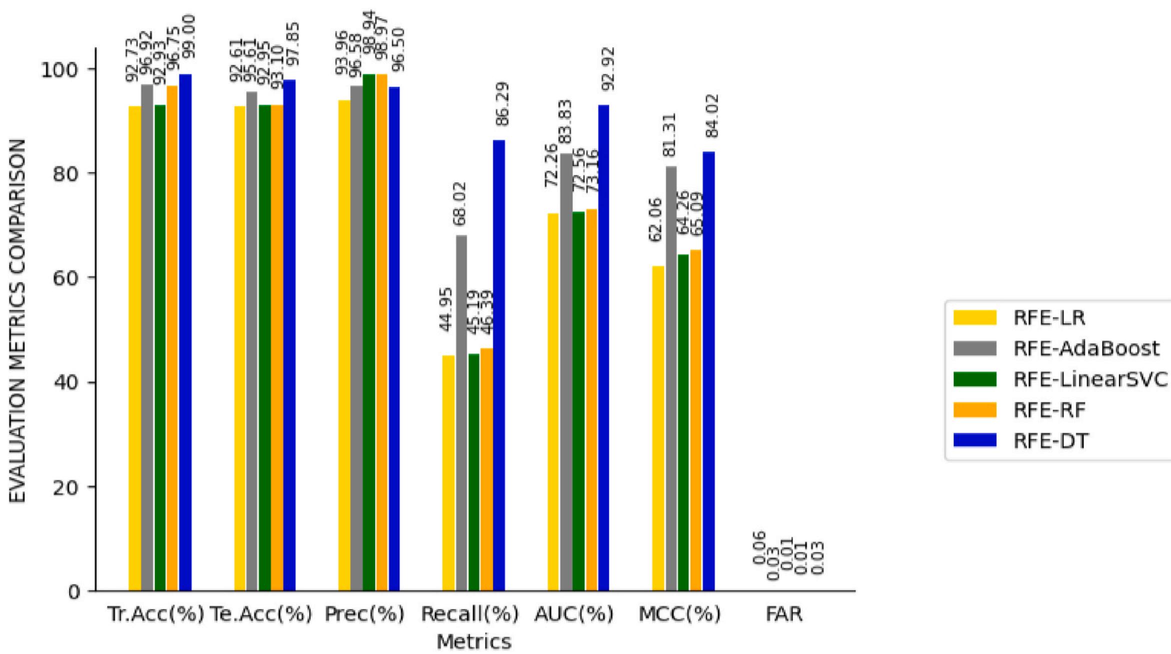

Fig. 6. Comparison of ML models metrics.

5.2. *Performance analysis of ridge-based DL models*

In this section. the performance detection of CNN and LSTM models was evaluated based on Ridge regression. Table 6 exhibits the evaluation results of the proposed Deep learning paradigms. According to results the Ridge-LSTM produced about 92.83 % training accuracy, 92.73 % testing accuracy, 95.43 % precision, 45.19 % recall, 72.43 % AUC, 62.84 % MCC, and 0.04 FAR. On another hand, the Ridge–CNN–based model performed a training accuracy of 92.59 %, and 92.34 % for testing accuracy. Precision, recall, and AUC of 89.52 %, 45.19 %, and 72.20 % respectively. Meanwhile, the observed metrics for the MCC and FAR are 60.37 %, and 0.10. Notably, in the case of DL models, the Ridge-LSTM produced promising results with an enhancement of 5.91 % in precision, and a marginal reduction of up to 0.06 in FAR compared to Ridge-CNN in binary classification. The use of the Ridge-CNN model in intrusion classification results in generalizing low-performance detection. Compared to LSTM which is capable of capturing time series data (Network traffic and patient biometrics data) for intrusion detection and classification, due to its capability to manage the temporal dependencies and patterns present in the data. It is evident from Fig. 7 that for performance evaluation metrics, the purple plot of Ridge-LSTM outshines the orange bar of Ridge-CNN, slightly in terms of training, testing accuracy, AUC, and MCC, however with a remarkable superiority in precision and FAR.

5.3. *Comparison of ML models vs DL paradigms*

The ML/DL models are compared by observing various performance metrics, moreover, Table 7 shows that all machine and deep learning models achieve a training accuracy range from 92 % to 99 %, testing accuracy between 92 % and 98 %, precision within 89 % and 99 %, recall inside of 44 %–86 %, AUC between 72 % and 93 %, MCC range from 60 % to 84 % and a FAR from 0.01 to 0.1. It can be noted that the RFE-ML models outperform Ridge-DL ones in terms of the seven aforementioned metrics, except LSTM which outweighs LR. Primarily due to the eight selected features per model performed by RFE, which helps to eliminate more noise and unnecessary features compared to the 19 ones chosen by Ridge. Specifically, the RFE-DT model realizes better results of 99 %, 97.85 %, 96.50 %, 86.29 %, 92.92 %, 84.02 %, and 0.03 respectively in training accuracy, testing accuracy, precision, recall, AUC, MCC, and FAR among the other machine learning and deep learning models, which means that the features selected by RFE-DT as seen in Table 4 are the more relevant for the intrusion detection task, contributing to its superior performance. Fig. 8 illustrates that the RFE-decision tree blue plot outperforms all other machine and deep learning model plots, while the RFE-DT precision still needs to be improved.

Table 6
Performance results for DL models.

Models	Training.acc	Testing.acc	Precision	Recall	AUC	MCC	FAR
Ridge-LSTM	92.83 %	92.73 %	95.43 %	45.19 %	72.43 %	62.84 %	0.04
Ridge-CNN	92.59 %	92.34 %	89.52 %	45.19 %	72.20 %	60.37 %	0.10

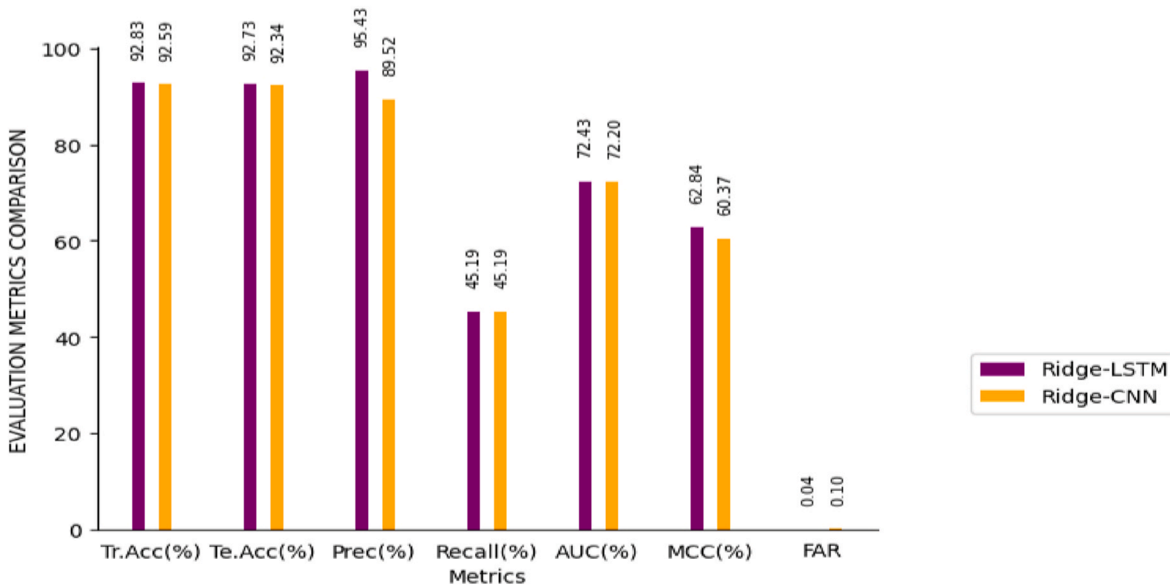

Fig. 7. Comparison of DL models performance.

Table 7
ML vs. DL models performance.

Models	Training.acc	Testing.acc	Precision	Recall	AUC	MCC	FAR
RFE-LR	92.73 %	92.61 %	93.96 %	44.95 %	72.26 %	62.06 %	0.06
RFE-AdaBoost	96.92 %	95.61 %	96.58 %	68.02 %	83.83 %	81.31 %	0.03
RFE-LinearSVC	92.93 %	92.95 %	98.94 %	45.19 %	72.56 %	64.26 %	0.01
RFE-RF	96.75 %	93.10 %	98.97 %	46.39 %	73.16 %	65.09 %	0.01
RFE-DT	**99 %**	**97.85 %**	**96.50 %**	**86.29 %**	**92.92 %**	**84.02 %**	**0.03**
Ridge-LSTM	92.83 %	92.73 %	95.43 %	45.19 %	72.43 %	62.84 %	0.04
Ridge-CNN	92.59 %	92.34 %	89.52 %	45.19 %	72.20 %	60.37 %	0.10

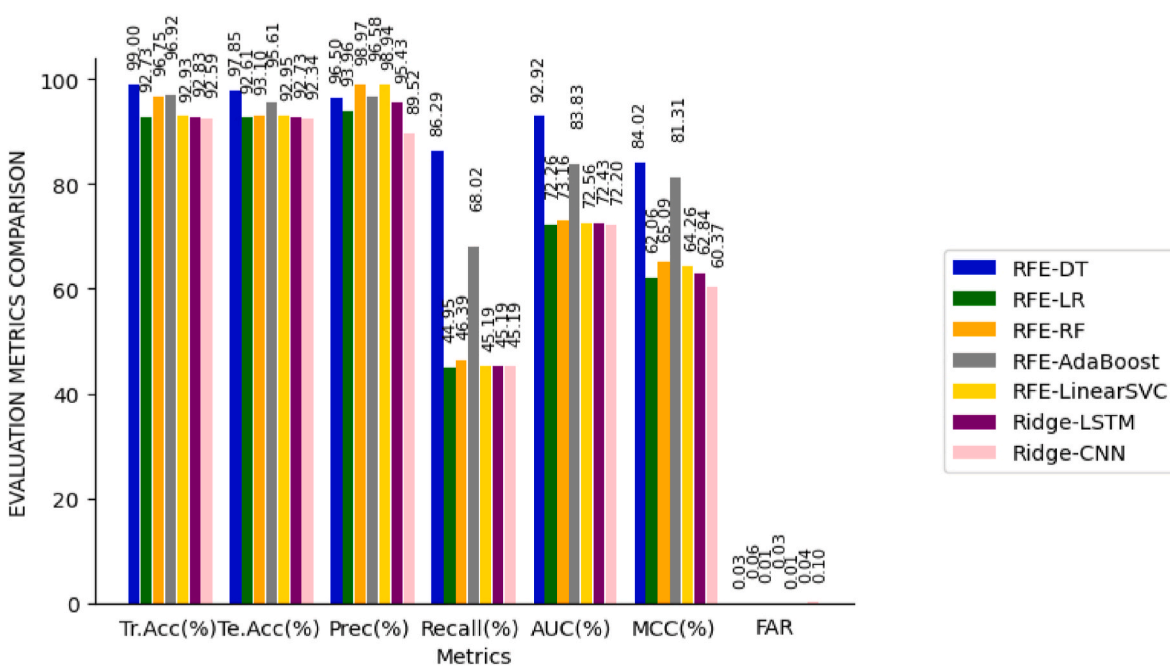

Fig. 8. Evaluation metrics for ML/DL models.

5.4. Memory usage analysis for feature selection and anomaly detection

In this section, the memory consumption of feature selection (F.S) using RFE and Ridge, and anomaly detection are evaluated for machine learning and deep learning models. As shown in Table 8, the process of feature selection in the RFE-Decision tree consumes 1.09 MB. Whereas

Table 8
Memory consumption during RFE/Ridge feature selection.

Feature selection memory	RFE-DT	RFE-LR	RFE-RF	RFE-LinearSVC	RFE-AdaBoost	Ridge-LSTM	Ridge-CNN
Memory usage	1.09 MB	10.74 MB	1.27 MB	1.70 MB	2.19 MB	2.52 MB	2.48 MB

10.74 MB for RFE-logistic regression, Additionally, the selection process in RFE-RF, RFE-SVC, and RFE-AdaBoost respectively expends 1.27 MB, 1.70 MB, and 2.19 MB. In contrast, usage memory of feature selection for Ridge-LSTM and Ridge-CNN obtains 2.52 MB and 2.48 MB. Table 9 summarizes the memory consumption of RFE-ML and Ridge-DL during anomaly detection (Grid search + ML/DL models). DT, LR, RF, SVC, and AdaBoost models consume 0.41 MB, 10.05 MB, 0.67 MB, 0.30 MB, and 1.51 MB respectively. However, LSTM and CNN depleted 38.18 MB and 12.34 MB.

The above results, show that in the case of RFE-ML models, the RFE feature selection process consumes more memory compared with anomaly detection, which is due to the fact of repetitive calculations required by RFE to train the model repeatedly by eliminating one or more characteristics at each iteration. These repetitive calculations increase memory consumption. In another hand, Ridge-DL proved a high memory usage in anomaly detection contrasted to Ridge feature selection because CNN requires convolutional operations that involve intensive matrix calculations. Furthermore, the batch size of CNN and LSTM is fixed to 64 which requires more memory to store multiple examples simultaneously.

As shown in Fig. 9 the feature selection in RFE-DT achieves the lowest memory consumption, the F.S in Ridge-CNN, and Ridge-LSTM nearly consumes the same memory, whereas F.S in RFE-Logistic regression expends the highest value. Then, we observe in Fig. 10 that RFE-DT, RFE-RF, and RFE-linearSVC realize approximately the same memory usage during anomaly detection with a slight difference, in contrast to Ridge-LSTM which wastes the greatest memory. The memory consumption of different models varies based on their configurations, hyperparameters, and the hardware used for experimentation. It is important to note that RFE-DT demonstrates the lowest memory consumption during the feature selection process, whereas RFE-linearSVC incurs the smallest memory usage when performing anomaly detection. Therefore, considering the detection performance, and the proposed intrusion detection task where accuracy and reliability are essential, the proposed RFE-DT model is the most efficient in terms of memory usage.

5.5. Comparison with the existing work

The choice of comparison papers is based on the classification category (normal or attack), and which dataset is used for evaluation. The papers mentioned in Section.4.3 are used as the benchmark for the results. Table 10 presents a detailed comparison of the recommended RFE-DT model with existing works, highlighting the observed results. Reference [24] suggested a tree classifier composed of random forest, robust scaled and data augmentation, their proposed model produced a training accuracy (Train.acc) of 92.85 %, testing accuracy (Test.acc) of 94.23 %, precision of 93.72 %, recall of 90.86 %, AUC of 90.68 %, and FAR of 0.06. The work [5] showcases a testing accuracy (Test.acc) of 96 %, precision of 96 %, and recall of 96 %, and did not focus on the two metrics MCC and FAR. Furthermore, based on the WUSTL-EHMS dataset, the proposed RFE-MLP framework in Ref. [39] realizes an accuracy of 96.20 %, a precision of 96.19 %, and a recall of 96.19 %. Additionally, the LRGU-MIFS system with DT classifier in Ref. [37] generated the highest accuracy average around 93.96 %. Using the CNN-ACL investigation, the authors in Ref. [38] reported on the KDDCUP99 dataset, an accuracy of 90.89 %, precision of 91 %, and 86 % for recall. On the CICIDS2017 dataset, they noted an accuracy of 91.74 %, a precision of 87 %, and a recall of 86 %. Moreover, the SmartHealth framework employed by Ref. [34] demonstrated an accuracy of 92 %, precision of 91 %, and recall of 92 %. Emphasizing our innovative approach, the proposed RFE-Decision tree model achieves the greatest training accuracy (Train.acc) of 99 %, testing accuracy (Test.acc) of 97.85 %, precision of 96.50 %, AUC of 92.92 %, MCC and FAR of 84.02 % and 0.03 over the related works. Whereas the recall value needs to be improved. A worthwhile analysis of Fig. 11 shows that the best-proposed model: RFE-DT model plotted in orange achieves high training, testing accuracy, precision, FAR, and AUC compared to the blue, red, purple, brown, and grey plots of the tree classifier model, RFE-MLP, LRGU-MIFS system, CNN-ACL-D2 on CICIDS2017 dataset, and SmartHealth respectively. Further, it surpasses the CNN-ACL-D1 plot on KDDCUP99 dataset, and the green plot tailored for PSO-DNN model which witnesses the absence of FAR. Taking into account the unavailability of MCC bars to evaluate the quality of binary classification model in the two aforementioned baselines.

On the same dataset, using RFE as a data analytic step plays a crucial part in enhancing the performance compared to the PSO, since RFE iteratively eliminates less important features, allowing more discriminant information to be preserved in the final model. By removing unimportant features, RFE can reduce Dataset dimensionality which enhances the effectiveness of machine learning algorithms. Meanwhile, Data Augmentation can increase the size of the dataset by adding new instances with transformations, which can introduce noise, potentially dilute discriminant information, and adversely affect model performance compared to the RFE which focuses on selecting the most essential features. Moreover, the Data augmentation produces overfitting which can lead to poor performance in contrast to RFE. Based on the proposed model results, the recall is the only potential limitation of the proposed work that must be improved. This can be achieved through the implementation of a voting classifier or stacking ensemble learning where multiple models are combined to improve overall predictive performance. Additionally, integrating active learning can further enhance recall by focusing on the most informative and challenging samples for labeling and training.

6. Conclusion

The proposed framework uses a wrapper and embedded feature selection techniques to select the most pertinent features and reduce dataset dimensionality as well as control the model complexity to encourage the paradigm to focus on the most informative features, to improve the model's effectiveness. The findings indicate that among all ML/DL models used, the best DT-RFE model reported an accuracy of 97.85 %, precision of 96.50 %, AUC of 92.92 %, MCC of 84.02 % and FAR of 0.03, in addition to efficient and modest memory consumption. The performance of the best model has been analyzed with the other

Table 9
Memory consumption during anomaly detection.

Feature selection memory	RFE-DT	RFE-LR	RFE-RF	RFE-LinearSVC	RFE-AdaBoost	Ridge-LSTM	Ridge-CNN
Memory usage	0.41 MB	10.05 MB	0.67 MB	0.30 MB	1.51 MB	38.18 MB	12.34 MB

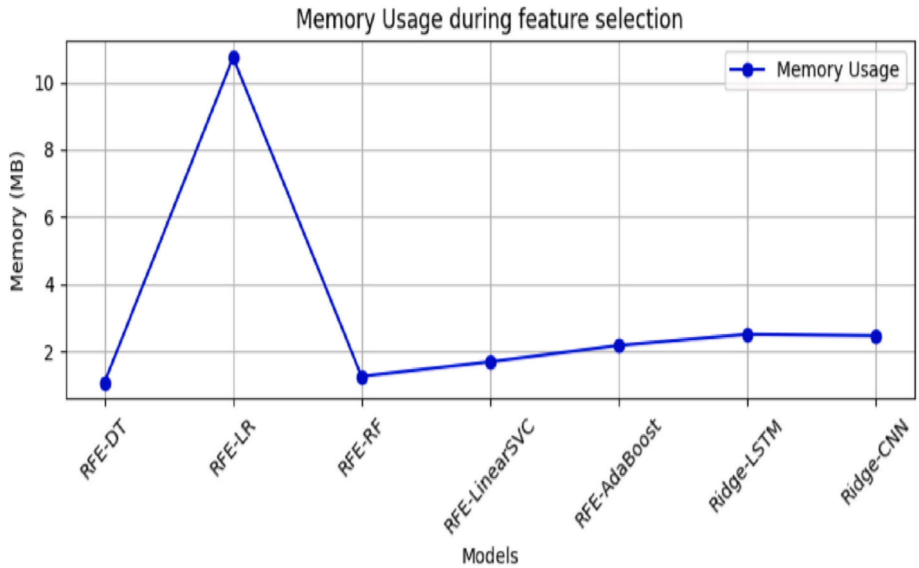

Fig. 9. Memory usage variation during feature selection.

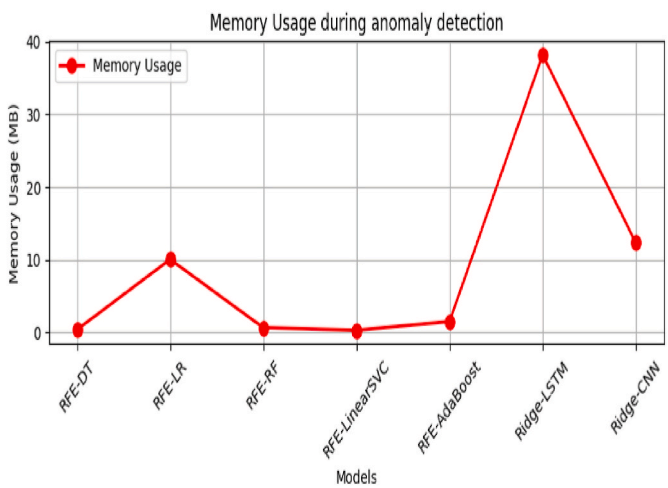

Fig. 10. Memory usage variation during anomaly detection.

state of-arts-works. The results also deduce that there is a meaningful augmentation in terms of the aforementioned metrics. In a nutshell, the results of this study provide insights into the implementation of RFE-RIDGE-based ML-DL models to create a viable IDS for the IoMT environment. In this regard, this study can serve as an initial step to mitigate the overbearing of the security mechanisms in IoMT hardware, and noteworthy to improve the AIDS performance for evolving cyber-attacks.

This work also has a scope to enhance the effectiveness and real-world applicability of our proposed security scheme, by exploring the integration of ML/DL and blockchain while using classical encryption techniques, which can be explored in future work. Investigating transfer learning techniques to adapt these pre-trained ML-DL models from the IoMT application domain to another IoT, like securing connected vehicles from various types of cyber threats, such as Distributed Denial of Service (DDoS) attacks, ransomware, and phishing attempts are considered essential. By leveraging knowledge gained from IoMT datasets and models, researchers can accelerate the deployment of robust anomaly detection systems in diverse IoT ecosystems. Additionally, we aim to test our approach on other datasets to further validate its effectiveness and adaptability across different scenarios.

Deploying and testing this approach within operational healthcare environments is crucial for validating its performance under real-world conditions. Healthcare institutions can provide access to live data streams, allowing researchers to refine the models based on actual cybersecurity incidents and challenges encountered in daily healthcare operations. However, working closely with healthcare institutions ensures that cybersecurity measures align with industry standards, regulatory requirements (such as HIPAA in the United States), and privacy laws governing patient data protection. Validating the models in compliance-sensitive environments enhances trust and facilitates adoption across the healthcare sector.

CRediT authorship contribution statement

Ghita Lazrek: Writing-original first draft, Investigation. **Kaouthar Chetioui:** Conceptualization, Methodology. **Younes Balboul:** Supervision. **Said Mazer & Moulhime EL Bekkali:**Expert views.

Table 10
Comparison with the existing studies.

Articles	Methods	Train.acc	Test.acc	Precision	Recall	AUC	MCC	FAR
[24]	Tree classifier	92.85 %	94.23 %	93.72 %	90.86 %	90.68 %	–	0.06
[5]	PSO-DNN	–	96 %	96 %	96 %	–	–	–
[39]	RFE-MLP	–	96.20 %	96.19 %	96.19 %	–	–	–
[37]	LRGU-MIFS	–	93.96 %	–	–	–	–	–
[38]	CNN-ACL	–	90.89 %	91 %	86 %	–	–	–
			91.47 %	87 %	86 %			
[34]	SmartHealth	–	92 %	91 %	92 %	–	–	–
This work	RFE-DT proposed model	99 %	97.85 %	96.50 %	86.29 %	92.92 %	84.02 %	0.03

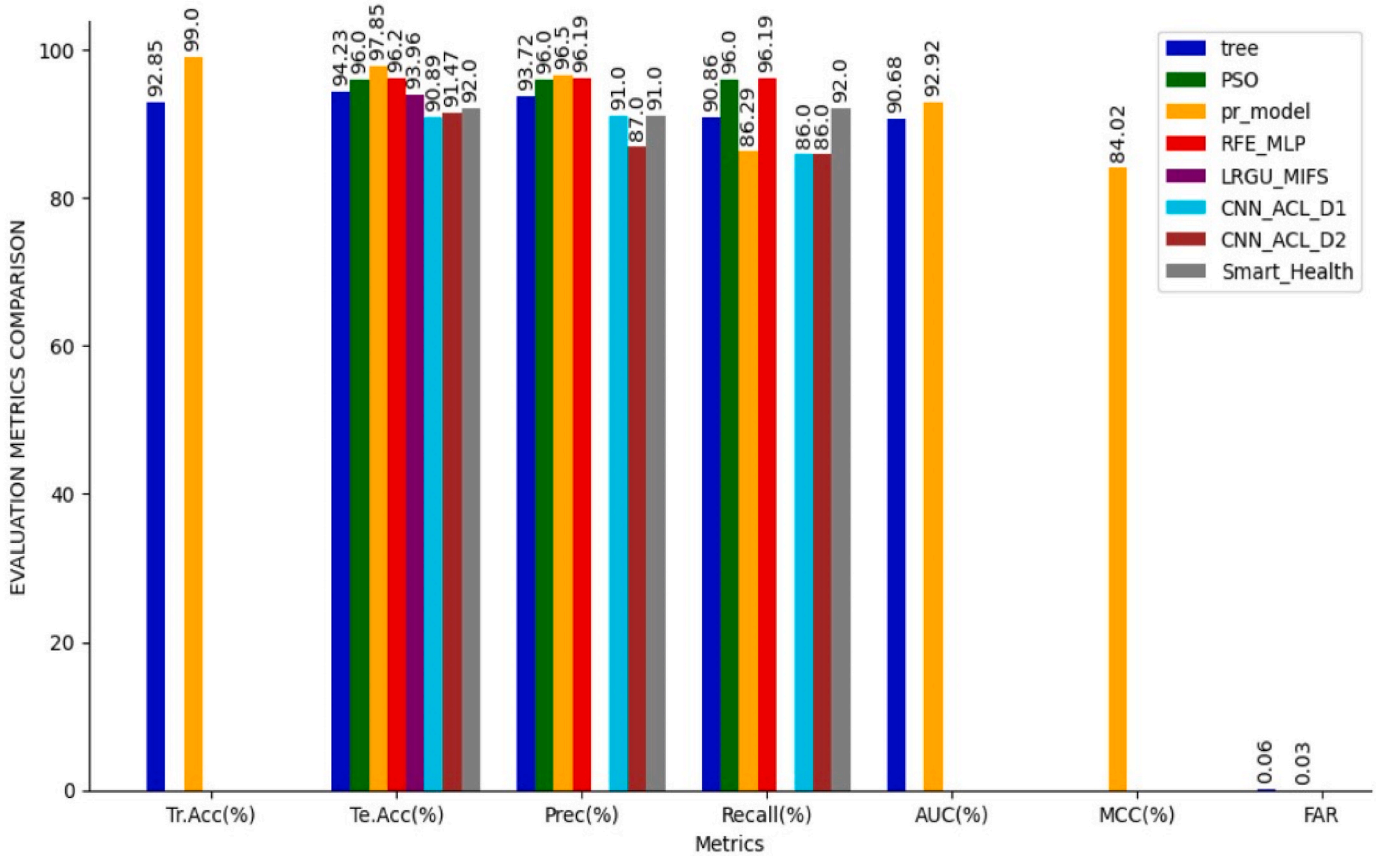

Fig. 11. Performance of the proposed model v/s existing models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] P.K. Sadhu, V.P. Yanambaka, A. Abdelgawad, Physical unclonable function and machine learning based group authentication and data masking for In-Hospital segments, Electronics 11 (24) (2022) 4155, https://doi.org/10.3390/electronics11244155.

[2] P. B, P. P, TwI-FTM: two-way IoT-FoG trust management scheme for task offloading in IoT-FoG networks, Results in Engineering 102197 (2024), https://doi.org/10.1016/j.rineng.2024.102197.

[3] Y.K. Saheed, M.O. Arowolo, Efficient cyber attack detection on the internet of medical Things-Smart environment based on deep recurrent neural network and machine learning algorithms, IEEE Access 9 (2021) 161546–161554, https://doi.org/10.1109/access.2021.3128837.

[4] N. Shingari, B. Mago, A framework for application-centric Internet of Things authentication, Results in Engineering 22 (2024) 102109, https://doi.org/10.1016/j.rineng.2024.102109.

[5] R. Chaganti, A. Mourade, V. Ravi, N. Vemprala, A. Dua, B. Bhushan, A particle swarm optimization and deep learning approach for intrusion detection system in internet of medical things, Sustainability 14 (19) (2022) 12828, https://doi.org/10.3390/su141912828.

[6] L. Singh, M. Kanstrup, K. Depa, A. Falk, V. Lindström, O. Dahl, K.E. Göransson, A. Rudman, E.A. Holmes, Digitalizing a brief intervention to reduce intrusive memories of psychological trauma for health care staff working during COVID-19: exploratory pilot study with nurses, JMIR Formative Research 5 (5) (2021) e27473, https://doi.org/10.2196/27473.

[7] A. Ghubaish, T. Salman, M. Zolanvari, D. Unal, A. Al-Ali, R. Jain, Recent advances in the Internet-of-Medical-Things (IOMT) systems security, IEEE Internet Things J. 8 (11) (2021) 8707–8718, https://doi.org/10.1109/jiot.2020.3045653.

[8] Q.A. Al-Haija, Cost-effective detection system of cross-site scripting attacks using hybrid learning approach, Results in Engineering 19 (2023) 101266, https://doi.org/10.1016/j.rineng.2023.101266.

[9] I.A. Khan, N. Moustafa, I. Razzak, M. Tanveer, D. Pi, Y. Pan, B.S. Ali, XSRU-IoMT: explainable simple recurrent units for threat detection in Internet of Medical Things networks, Future Generat. Comput. Syst. 127 (2022) 181–193, https://doi.org/10.1016/j.future.2021.09.010.

[10] W. Hurst, B. Tekinerdogan, T. Alskaif, A. Boddy, N. Shone, Securing electronic health records against insider-threats: a supervised machine learning approach, Smart Health 26 (2022) 100354, https://doi.org/10.1016/j.smhl.2022.100354.

[11] F. Khan, M.A. Jan, R. Alturki, M.D. Alshehri, S.T. Shah, A.U. Rehman, A secure ensemble learning-based fog-cloud approach for cyberattack detection in IOMT, IEEE Trans. Ind. Inf. 19 (10) (2023) 10125–10132, https://doi.org/10.1109/tii.2022.3231424.

[12] D. Noori, H. Shakeri, M.N. Torshiz, Scalable, efficient, and secure RFID with elliptic curve cryptosystem for Internet of Things in healthcare environment, EURASIP J. Inf. Secur. 2020 (1) (2020), https://doi.org/10.1186/s13635-020-00114-x.

[13] J. Kang, K. Fan, K. Zhang, X. Cheng, H. Li, Y. Yang, An ultra light weight and secure RFID batch authentication scheme for IoMT, Comput. Commun. 167 (2021) 48–54, https://doi.org/10.1016/j.comcom.2020.12.004.

[14] K. Sowjanya, M. Dasgupta, S. Ray, Elliptic curve Cryptography based authentication scheme for internet of medical things, J. Inf. Secur. Appl. 58 (2021) 102761, https://doi.org/10.1016/j.jisa.2021.102761.

[15] M.M. Salim, I. Kim, U. Doniyor, C. Lee, J.H. Park, Homomorphic encryption based privacy-preservation for IOMT, Appl. Sci. 11 (18) (2021) 8757, https://doi.org/10.3390/app11188757.

[16] Y. Sun, F.P. Lo, B. Lo, Security and privacy for the Internet of medical Things enabled healthcare Systems: a survey, IEEE Access 7 (2019) 183339–183355, https://doi.org/10.1109/access.2019.2960617.

[17] S. Manimurugan, S. Al-Mutairi, M.M. Aborokbah, N. Chilamkurti, S. Ganesan, R. Patan, Effective attack detection in internet of medical things smart environment using a deep belief neural network, IEEE Access 8 (2020) 77396–77404, https://doi.org/10.1109/access.2020.2986013.

[18] S.P. Rm, P.K.R. Maddikunta, P. M, S. Koppu, T.R. Gadekallu, C.L. Chowdhary, M. Alazab, An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture, Comput. Commun. 160 (2020) 139–149, https://doi.org/10.1016/j.comcom.2020.05.048.

[19] P. Kumar, G.P. Gupta, R. Tripathi, An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks, Comput. Commun. 166 (2021) 110–124, https://doi.org/10.1016/j.comcom.2020.12.003.

[20] C. Iwendi, J.H. Anajemba, C. Biamba, D. Ngabo, Security of things intrusion detection system for smart healthcare, Electronics 10 (12) (2021) 1375, https://doi.org/10.3390/electronics10121375.

[21] G. Zachos, I. Essop, G. Mantas, K. Porfyrakis, J.C. Ribeiro, J. Rodriguez, An Anomaly-Based intrusion detection system for internet of medical things networks, Electronics 10 (21) (2021) 2562, https://doi.org/10.3390/electronics10212562.

[22] S. Khan, A. Akhunzada, A hybrid DL-driven intelligent SDN-enabled malware detection framework for Internet of Medical Things (IoMT), Comput. Commun. 170 (2021) 209–216, https://doi.org/10.1016/j.comcom.2021.01.013.

[23] A. Binbusayyis, H. Alaskar, T. Vaiyapuri, M. Dinesh, An investigation and comparison of machine learning approaches for intrusion detection in IoMT network, ~the œJournal of Supercomputing/Journal of Supercomputing 78 (15) (2022) 17403–17422, https://doi.org/10.1007/s11227-022-04568-3.

[24] K. Gupta, D.K. Sharma, K.D. Gupta, A. Kumar, A tree classifier based network intrusion detection model for Internet of Medical Things, Comput. Electr. Eng. 102 (2022) 108158, https://doi.org/10.1016/j.compeleceng.2022.108158.

[25] S. Saif, P. Das, S. Biswas, M. Khari, V. Shanmuganathan, HIIDS: hybrid intelligent intrusion detection system empowered with machine learning and metaheuristic algorithms for application in IoT based healthcare, Microprocess. Microsyst. 104622 (2022), https://doi.org/10.1016/j.micpro.2022.104622.

[26] M. Zubair, A. Ghubaish, D. Unal, A. Al-Ali, T. Reimann, G. Alinier, M. Hammoudeh, J. Qadir, Secure Bluetooth communication in smart healthcare systems: a novel community dataset and intrusion detection system, Sensors 22 (21) (2022) 8280, https://doi.org/10.3390/s22218280.

[27] N.I. Haque, M.A. Rahman, S.I. Ahamed, DeepCAD: a stand-alone deep neural network-based framework for classification and anomaly detection in smart healthcare systems. 2022 IEEE International Conference on Digital Health (ICDH), 2022, https://doi.org/10.1109/icdh55609.2022.00042.

[28] S. Nandy, M. Adhikari, M.A. Khan, V.G. Menon, S. Verma, An intrusion detection mechanism for secured IOMT framework based on Swarm-Neural network, IEEE Journal of Biomedical and Health Informatics 26 (5) (2022) 1969–1976, https://doi.org/10.1109/jbhi.2021.3101686.

[29] J.B. Awotunde, K.M. Abiodun, E.A. Adeniyi, S.O. Folorunso, R.G. Jimoh, A Deep Learning-Based intrusion detection technique for a secured IOMT system, in: Communications in Computer and Information Science, 2022, pp. 50–62, https://doi.org/10.1007/978-3-030-95630-1_4.

[30] S. Abbas, G.A. Sampedro, M. Abisado, A. Almadhor, I. Yousaf, S. Hong, Harris-Hawk-Optimization-Based deep recurrent neural network for securing the internet of medical things, Electronics 12 (12) (2023) 2612, https://doi.org/10.3390/electronics12122612.

[31] S. Saif, N. Yasmin, S. Biswas, Feature engineering based performance analysis of ML and DL algorithms for Botnet attack detection in IoMT, International Journal of System Assurance Engineering and Management 14 (S1) (2023) 512–522, https://doi.org/10.1007/s13198-023-01883-7.

[32] N. Goswami, S. Raj, D. Thakral, J.L. Arias-Gonzáles, J. Flores-Albornoz, E. Asnate-Salazar, D. Kapila, S. Yadav, S. Kumar, Intrusion detection system for IoT-based healthcare intrusions with lion-salp-swarm-optimization algorithm: metaheuristic-enabled hybrid intelligent approach, Engineered Science (2023), https://doi.org/10.30919/es933.

[33] A.a.J. Al-Abadi, M.B. Mohamed, A. Fakhfakh, Robust and reliable security approach for IoMT: detection of DoS and delay attacks through a high-accuracy machine learning model, International Journal on Recent and Innovation Trends in Computing and Communication 11 (6) (2023) 239–247, https://doi.org/10.17762/ijritcc.v11i6.7558.

[34] S. Rani, S. Kumar, A. Kataria, H. Min, SmartHealth: an intelligent framework to secure IoMT service applications using machine learning, ICT Express (2023), https://doi.org/10.1016/j.icte.2023.10.001.

[35] N. Faruqui, M.A. Yousuf, M. Whaiduzzaman, A. Azad, S.A. Alyami, P. Liò, M. A. Kabir, M.A. Moni, SafetyMed: a novel IOMT intrusion detection system using CNN-LSTM hybridization, Electronics 12 (17) (2023) 3541, https://doi.org/10.3390/electronics12173541.

[36] P. Kulshrestha, T.V.V. Kumar, Machine learning based intrusion detection system for IoMT, International Journal of System Assurance Engineering and Management (2023), https://doi.org/10.1007/s13198-023-02119-4.

[37] M. Alalhareth, S. Hong, An improved mutual information feature selection technique for intrusion detection systems in the internet of medical things, Sensors 23 (10) (2023) 4971, https://doi.org/10.3390/s23104971.

[38] J. Mathew, R.J. Priyadarsini, Efficient DDOS detection in internet of medical things using CNN-ACL approach, International Journal of Intelligent Systems and Applications in Engineering 11 (4) (2023) 789–799. Retrieved from, https://ijisae.org/index.php/IJISAE/article/view/3612.

[39] I.F. Kilincer, F. Ertam, A. Sengur, R. Tan, U.R. Acharya, Automated detection of cybersecurity attacks in healthcare systems with recursive feature elimination and multilayer perceptron optimization, Biocybern. Biomed. Eng. 43 (1) (2023) 30–41, https://doi.org/10.1016/j.bbe.2022.11.005.

[40] P.G. Shambharkar, N. Sharma, Deep learning-empowered intrusion detection framework for the Internet of Medical Things environment, Knowl. Inf. Syst. (2024), https://doi.org/10.1007/s10115-024-02149-9.

[41] U. Zukaib, X. Cui, C. Zheng, M. Hassan, Z. Shen, Meta-IDS: meta-learning based smart intrusion detection system for internet of medical things (IOMT) network, IEEE Internet Things J. 1 (2024), https://doi.org/10.1109/jiot.2024.3387294.

[42] N. Tendikov, L. Rzayeva, B. Saoud, I. Shayea, M.H. Azmi, A. Myrzatay, M. Alnakhli, Security information event Management data acquisition and analysis methods with machine learning principles, Results in Engineering 102254 (2024), https://doi.org/10.1016/j.rineng.2024.102254.

[43] G. Lazrek, K. Chetioui, Y. Balboul, Enhancing IOMT Security: a conception of RFE-Ridge and ML/DL for anomaly intrusion detection, in: Lecture Notes in Networks and Systems, 2024, pp. 442–447, https://doi.org/10.1007/978-3-031-48573-2_63.

[44] GeeksforGeeks, ML Overview of data cleaning. GeeksforGeeks. https://www.geeksforgeeks.org/data-cleansing-introduction/, 2024 accessed on 18 June 2023.

[45] GeeksforGeeks, Label Encoding in Python, GeeksforGeeks, 2023. https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/. accessed on 15 June 2023.

[46] D. Radečić, Feature selection in Python — recursive feature elimination, Medium (2022). https://towardsdatascience.com/feature-selection-in-python-recursive-feature-elimination-19f1c39b8d15. accessed on June 24, 2023.

[47] A hands-on guide to ridge regression for feature selection. https://analyticsindiamag.com/developers-corner/a-hands-on-guide-to-ridge-regression-for-feature-selection/, 2022 accessed on 26 June 2023.

[48] K. Yasar, G. Lawton, E. Burns, Logistic regression, Business Analytics, https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression, 2024. accessed on 28 June 2023.

[49] Speech and Language Processing. (n.d.). https://web.stanford.edu/~jurafsky/slp3/(last accessed 5 January 2024).

[50] What is boosting and AdaBoost in machine learning?. https://www.knowledgehut.com/blog/data-science/boosting-and-adaboost-in-machine-learning, 2024 accessed on 28 June 2023.

[51] N. Donges, Random Forest: a complete guide for machine learning, Built In (2024). https://builtin.com/data-science/random-forest-algorithm. accessed on 25 June 2023.

[52] A. Arshad, M. Jabeen, S. Ubaid, A. Raza, L. Abualigah, K. Aldiabat, H. Jia, A novel ensemble method for enhancing Internet of Things device security against botnet attacks, Decision Analytics Journal 8 (2023) 100307, https://doi.org/10.1016/j.dajour.2023.100307.

[53] H. Al-Manaseer, L. Abualigah, A.R. Alsoud, R.A. Zitar, A. Ezugwu, H. Jia, A novel big data classification technique for healthcare application using support Vector Machine, Random Forest and J48, in: Studies in Computational Intelligence, 2022, pp. 205–215, https://doi.org/10.1007/978-3-031-17576-3_9.

[54] Decision Tree Classification in Python Tutorial. (n.d.). https://www.datacamp.com/tutorial/decision-tree-classification-python (last accessed 26 June 2023).

[55] A.H. Gandomi, F. Chen, L. Abualigah, Machine learning technologies for big data analytics, Electronics 11 (3) (2022) 421, https://doi.org/10.3390/electronics11030421.

[56] F.W. Alsaade, M.H. Al-Adhaileh, Cyber attack detection for Self-Driving vehicle networks using deep autoencoder algorithms, Sensors 23 (8) (2023) 4086, https://doi.org/10.3390/s23084086.

[57] J. Cho, S. Gong, K. Choi, A study on High-Speed Outlier Detection method of network abnormal behavior data using heterogeneous multiple classifiers, Appl. Sci. 12 (3) (2022) 1011, https://doi.org/10.3390/app12031011.

[58] A. Kowalczyk, *Svm - Understanding the math : the optimal hyperplane*. SVM Tutorial. https://www.svm-tutorial.com/2015/06/svm-understanding-math-part-3/, 2023 accessed on 2 July 2023.

[59] V. Choubey, Text classification using CNN - voice tech podcast - medium, Medium (2023). https://medium.com/voice-tech-podcast/text-classification-using-cnn-9ade8155dfb9. accessed on 29 June 2023.

[60] S. Verma, Understanding 1D and 3D convolution neural network | keras, Medium (2023). https://towardsdatascience.com/understanding-1d-and-3d-convolution-neural-network-keras-9d8f76e29610. accessed on 3 July 2023.

[61] A Gentle Introduction to Autocorrelation and Partial Autocorrelation. (n.d.). https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/(last accessed 30 June 2023).

[62] S. Saxena, What is LSTM? Introduction to long short-term memory. Analytics Vidhya, 2024. https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/. accessed on 31 June 2023.

[63] S. Al, M. Dener, STL-HDL: a new hybrid network intrusion detection system for imbalanced dataset on big data environment, Comput. Secur. 110 (2021) 102435, https://doi.org/10.1016/j.cose.2021.102435.

[64] A.A. Hady, A. Ghubaish, T. Salman, D. Unal, R. Jain, Intrusion Detection System for healthcare systems using medical and network Data: a comparison study, IEEE Access 8 (2020) 106576–106584, https://doi.org/10.1109/access.2020.3000421.

[65] WUSTL EHMS 2020 Dataset for Internet of Medical Things (IoMT) Cybersecurity Research. (n.d.). https://www.cse.wustl.edu/~jain/ehms/index.html (last accessed 15 December 2024).

[66] Anaconda, Download anaconda distribution | Anaconda. https://www.anaconda.com/download/, 2024 accessed on 4 June 2023.

[67] Team, K. (n.d.). Keras: Deep Learning for humans. https://keras.io/(last accessed 30 May 2023).

[68] Nik, Hyper-parameter Tuning with GridSearchCV in Sklearn, Datagy, 2023. https://datagy.io/sklearn-gridsearchcv/. accessed on 21 June 2023.

[69] S. Loukas, Everything you need to know about Min-Max normalization: a Python tutorial, Medium (2024). https://towardsdatascience.com/everything-you-need-to-know-about-min-max-normalization-in-python-b79592732b79. accessed on 17 June 2023.

[70] A. Bajaj, Performance metrics in machine learning [complete guide], neptune.ai, https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide, 2023. accessed on 29 June 2023.