IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 29, 2022, accepted April 25, 2022, date of publication May 3, 2022, date of current version May 11, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3172432

A Security Model Based on LightGBM and Transformer to Protect Healthcare Systems From Cyberattacks

ABDALLAH GHOURABI
Department of Computer Science, College of Science and Arts, Jouf University, Sakaka, Al Jowf 72388, Saudi Arabia
e-mail: aghourabi@ju.edu.sa

This work was supported by the Deanship of Scientific Research at Jouf University under Grant DSR-2021-02-0108.

ABSTRACT Cybersecurity incidents have become a growing problem for the healthcare industry since the widespread introduction of technology into the healthcare systems. In recent years, the number of attacks has increased rapidly in healthcare, and it is now among the sectors most targeted by cyberattacks globally. These types of attacks are not only a threat to the data and finances of medical organizations, but they can also disrupt hospital operations and endanger the health and well-being of patients. Traditional security measures are not sufficient to protect the healthcare IT (Information Technology) environment due to its complexity and the heterogeneity of its medical devices. In this paper, we propose a new intrusion and malware detection system to secure the entire network of the healthcare system. The proposed solution includes two components: an intrusion detection system for medical devices installed in the healthcare network, and a malware detection system for data servers and medical staff computers. The objective is to secure the entire network independently of the installed devices and computers. The proposed system is based on an optimized LightGBM model and a Tranformer-based model. It is trained with four different datasets to guarantee a varied knowledge of the different types of attacks that can affect the healthcare sector. The used datasets have been generated from different environments undergoing IoT (Internet of Things), IoMT (Internet of Medical Things) and Windows malware attacks. The experimental evaluation of the approach showed remarkable accuracies of 99%.

INDEX TERMS Healthcare security, intrusion detection, malware detection, LightGBM, transformer.

I. INTRODUCTION

In recent years, the healthcare sector has experienced a large number of unprecedented cyberattacks [1]. This phenomenon is increased especially with the COVID-19 pandemic. According to reports from the World Health Organization (WHO) [2], there was a five-fold increase in the number of cyberattacks launched during the COVID-19 pandemic. These malicious operations are varied, they include: malwares that attempt to compromise the integrity of the systems; phishing operations; or DDoS attacks that seek to disrupt the ability of facilities to provide patient care. These types of attacks not only result in financial loss and privacy violation, but they can also disrupt hospital operations and place the health of patients at risk.

As an example, we cite the ‘‘WannaCry’’ ransomware attack in 2017, which invaded hundreds of thousands of computers in over 150 countries. This ransomware is a malicious software that encrypts the victim’s data and makes it inaccessible, and then asks to pay a ransom in order to recover that data. The most profound impact of WannaCry occurred in UK when the National Health Service (NHS) systems were infected. The result was a disruption of normal operations at more than 80 hospitals for four days, resulting in delays in patient care [3]. Most recently, in September 2020, a ransomware attack on the University Hospital of Düsseldorf in Germany resulted in a system and data access failure, which led the hospital to turn away patients in an emergency. During the attack, a woman who had to be sent to another health facility about 30 km away, died because of the delay in treatment [4].

Although the field of information and communication technology is very rich in terms of security techniques and

The associate editor coordinating the review of this manuscript and approving it for publication was Mounim A. El Yacoubi.

tools, these means are not appropriate for many medical devices that have different characteristics. Medical devices are very varied, their architecture differs from one device to another, their communication techniques and protocols are not the same. Processing and storage capacities are limited. All these constraints make it extremely difficult to implement practical security mechanisms for healthcare devices.

In the literature, research on the subject of security of healthcare systems is still limited. In this context, some research [5]–[8] focuses on the implementation of authentication and encryption solutions for implanted and wearable medical devices. Such solutions are generally computationally expensive and difficult to deploy on medical devices with limited resources. Other research works [9]–[12] propose intrusion and malware detection systems for IoMT (Internet of Medical Things) networks based on machine learning algorithms. These models are generally trained with standard datasets intended for traditional detection systems. It makes it difficult to identify certain attacks specific to medical devices. This is due to the lack of public datasets monitoring attacks carried out in healthcare environments. To the best of our knowledge, there are currently only two public datasets focusing on attacks targeting medical devices ECU-IoHT [13] and ICE [14]. Despite their benefit, in our opinion, these two datasets are insufficient to develop robust intrusion or malware detection systems. They were generated in limited IoMT environments simulating a few medical devices and experiencing some cyberattacks.

In contrast to existing solutions, we propose in this paper a more complete approach to protect healthcare systems against cyber-attacks. Our solution tries to take into account the complexity and heterogeneity of the healthcare network. The idea is to provide a hybrid security system composed of two components: the first one is an intrusion detection system to monitor the IoMT network containing the medical devices of the healthcare organization, the second one is a malware detection system to inspect the computers of the medical staff and check for malware. The goal is to provide an entire protection for the healthcare network regardless of the type of used device. The proposed solution is based on machine learning algorithms which allows to detect known and unknown attacks. Two machine learning models are used in the approach: LightGBM and BERT-based Transformer. The training of these models is performed with four different datasets: ECU-IoHT to analyze attacks targeting medical devices, ToN-IoT and Edge_IIoTset for IoT (Internet of Things) attacks, and EMBER to track malware on Windows environments. Our objective is to create a powerful healthcare security system that can deal with different types of attacks regardless of the targeted device.

The main contributions of our paper are summarized as follows:

- Proposal of a hybrid security solution capable of detecting different types of attacks in a healthcare environment.
- An intrusion detection system based on LightGBM and Transformer for IoHT and IoT attacks.
- A malware detection system based on LightGBM for healthcare staff computers.
- The proposed system outperforms existing solutions by achieving remarkable accuracies very close to 100%.

The remaining parts of the paper are organized as follows: Section 2 reviews the related works. Section 3 explains the machine learning techniques used in our approach. Section 4 describes the model design of our approach. Section 5 presents the experimental results. Finally, we conclude the paper in Section 6.

II. RELATED WORK

In the literature, different methods have been used to propose security solutions for healthcare systems. In this section, we grouped the related works into 3 categories: authentication and access control of medical data, intrusion detection systems, and malware detection systems.

A. AUTHENTICATION AND ACCESS CONTROL

In [5], the authors proposed a device to device (D2D) access control scheme called "D2DAC-IoMT" in order to secure device to device communication in Internet of medical things systems. The access control to devices is carried out using the symmetric key primitives and Elliptic-curve cryptography ECC based device specific certificates.

In [6], the authors proposed an approach to secure the communication channels of implantable medical devices against man-in-the-middle attacks. In the proposed approach, the data transmission is secured through using authenticated encryption with a random vector and a timestamp encoded by Robust or AMD codes. The authors also sought to minimize the power consumption of their system and introduced a secure protocol for authorizing third-party medical members to access the implantable medical devices in the case of an emergency.

In [7], Aledhari *et al.* used genomics encryption and deterministic chaos to design a cryptography algorithm in order to secure lightweight wearable medical devices. The objective of the proposed algorithm is to protect healthcare data from major threats, such as key theft, man-in-the-middle attack, and brute force attack.

In [8], the authors chose to use blockchain to create a decentralized authentication mechanism of patients and medical staff in a distributed hospital network. The principle of the proposed approach is to realize a blockchain-based P2P communication between the members of the distributed hospital network. It also allows the migration of authenticated participants from one affiliated hospital to another without re-authentication requirements.

In another paper, Garg *et al.* [15] designed an authentication key agreement protocol based on blockchain for IoMT environment. The proposed protocol provides secure key management between implantable medical devices and personal servers and between personal servers and cloud servers. According to the authors, the formal security verification shown that the protocol is robust against various known passive/active attacks.

B. INTRUSION DETECTION

In [9], the authors proposed an Intrusion Detection and Prevention System to protect the healthcare communications based on HTTP and Modbus/TCP. The choice of these two protocols is justified by the fact that the HTTP protocol is used by several healthcare applications, and the Modbus/TCP protocol is an industrial protocol that is adopted by several medical devices. To adopt their system to healthcare environments, the authors proposed an active learning approach to make it able to re-train itself. To evaluate the approach, several machine learning classifiers were tested including Random Forest, Decision Tree, SVM, KNN, Naive Bayes, MLP and DNN.

In a recent paper [13], the authors developed a dataset named ECU-IoHT designed to analyze attacks targeting the Internet of Health Things (IoHT). This dataset was created by performing several types of attacks targeting a healthcare environment containing devices such as temperature sensor, blood pressure sensor and heart rate sensor. To evaluate the effectiveness of the dataset, the authors designed an anomaly detection system using several machine learning algorithms such as k-Nearest Neighbor, Local Outlier Factor, Influenced Outlierness and SVM.

In [16], Hady *et al.* designed an Enhanced Healthcare Monitoring System (EHMS) testbed that monitors the patients' biometrics data and collects network flow metrics. This system helped them to collect a dataset of 16 thousand records of normal and attack healthcare data. On the dataset, they tested an intrusion detection system by comparing four ML algorithms, RF, KNN, SVM, and ANN.

In [17], the authors proposed an intrusion detection system that relies on data collected from real medical devices. The proposed system uses convolutional neural networks and classifies the detected intrusions into four classes: critical, informal, major, and minor.

In [10], the authors proposed a mobile agent-driven intrusion detection system for Internet of Medical Things. The system is dedicated to detect network level intrusions and device level anomalies. For the network level, the authors used machine learning algorithms, while for the device level they used polynomial regression.

C. MALWARE DETECTION

In [11], the authors proposed a hybrid deep learning model for the detection of malwares in Internet of Medical Things (IoMT). The proposed model is an SDN-enabled framework and is based on two deep learning algorithms, CNN and LSTM.

In [14], the authors presented an intelligent system capable of detecting, classifying and mitigating ransomware attacks affecting hospital rooms equipped with integrated clinical environments. The detection and the classification of ransomware attacks is based on network flow analysis of the clinical environment. The authors evaluated several machine learning algorithms and selected One-Class SVM for attack detection and Naive Bayes for ransomware classification. In addition, the authors proposed a protection method consisting of isolating infected medical devices through the SDN paradigm and replacing their software controllers using NFV techniques.

In recent paper [12], Anand *et al.* proposed a deep learning model based on CNN to detect malware attacks in 5G-IoT healthcare applications. The main idea of the model is to convert the input binary into gray-scale image and then apply a CNN classifier to detect malware.

In [18], the authors proposed a behavior-monitoring system to detect malware on embedded medical devices. The proposed system is based on monitoring the systemwide power consumption of the devices without modifying their embedded software. The idea is to monitor the behavior of the devices and detect any aberrant behavior indicating the presence of malware. To model and classify device behavior, the authors tested machine learning algorithms such as 3-nearest neighbors, multilayer perceptron, and random forest.

In Table 1, we present a comparative summary on different works discussed in this section. Despite the importance and added value of these works, they are not yet able to provide a global and complete protection of the healthcare network, which is known for its heterogeneity and the diversity of its devices. For example, authentication and access control solutions are very useful to protect the integrity of medical data, but they are not able to prevent disruptions in the operation of devices or malware attacks. As for intrusion and malware detection systems, their learning capabilities are limited to standard datasets without knowledge of the specific attacks on medical devices. While there are newer datasets for healthcare attacks such as ECU-IoHT or ICE, these have been generated in limited IoMT environments, which affects the ability to detect other types of attacks. In contrast to existing solutions, in this paper we propose a more complete approach to protect healthcare systems from cyberattacks. Our goal is to provide an intelligent intrusion and malware detection system trained with multiple datasets. This includes different types of attacks conducted in different environments, including IoMT networks, IoT networks, and Windows environments. This helps our system to get a broader view of the different attacks that can occur in a healthcare environment. In addition, we employ recent machine learning models in this system, which have resulted in state-of-the-art accuracy in several applications.

III. BACKGROUND

In this section, we provide background details regarding the machine learning methods used in this paper: BERT-based Transformer, LightGBM and BiLSTM.

A. BERT-BASED TRANSFORMER

Transformer is a deep learning model that relies on an attention mechanism to establish global dependencies between

TABLE 1. Comparative summary of related work.

Paper reference	Approach Objective	Used techniques	Target environment	Dataset type
[5]	Device to device access control	Elliptic-curve cryptography certificate	Internet of medical things	–
[6]	Securing data transmission of implantable medical devices	Authenticated encryption	Implantable medical devices	–
[7]	Secure wearable medical devices	Genomics encryption and deterministic chaos	Wearable medical devices	–
[8]	Decentralized authentication of patients in a distributed hospital network	Blockchain	patients, medical staff, IoT devices	–
[15]	Authentication key management protocol for IoMT	Blockchain	IoMT	–
[9]	Intrusion detection and prevention system for healthcare environment	Active learning + ML classifiers	Healthcare applications and devices	CIC-IDS2017 [19] + emulated data
[13]	Anomaly detection for healthcare environment	ML algorithms	Healthcare devices	ECU-IoHT [13]
[16]	Intrusion detection system for healthcare environment	ML algorithms	Healthcare devices	private dataset
[17]	Intrusion detection system for healthcare environment	Convolutional neural networks	Healthcare devices	Data collected from real medical devices
[10]	Intrusion detection system for IoMT	Machine learning and polynomial regression	Medical network and devices	Data collected from simulated network
[11]	Malware detection framework for Internet of Medical Things	CNN and LSTM	IoMT	IoT malware dataset
[14]	Detection, Classification and mitigation of ransomware attacks affecting clinical environments	SVM and Naïve Bayes	Integrated Clinical Environments	Clinical environments malware dataset [14]
[12]	Malware detection and classification	CNN	IoT healthcare applications	Malimg dataset [20]
[18]	Malware detection	ML algorithms	Embedded medical devices	private dataset

input and output [21]. Since their introduction in 2017, the transformer models have significantly improved the NLP area. Thus, several pre-trained models based on the Transformer architecture have appeared. BERT [22] was one of these models that had great success in solving several NLP problems. The BERT model is pre-trained on a large volume of data and can be fine-tuned with neural layers to perform NLP operations such as sentiment analysis, text classification, machine translation, etc.

The purpose of using BERT in the model proposed in this paper is to generate textual sequence from the input network flow and classify it as normal or attack. To clarify this idea, we describe in this paragraph how to fine-tune the pre-trained model for sequence classification. We show an example of fine-tuning BERT on single phrase classification in Figure 1. A sequence of tokens is used to represent each input sentence. Every sequence begins with a particular classification token that is noted ([CLS]). The input sequence embedding is denoted by E, the final hidden vector of the special [CLS] token is denoted by $C \in \mathbb{R}^H$, where H is the hidden layer size [21]. The classification task can be performed by connecting the final hidden vector of the [CLS] token to supplementary layers.

B. LightGBM

LightGBM (Light Gradient Boosting Machine) is an open-source distributed gradient boosting framework for machine learning that was created by Microsoft in 2017 [23]. It is based on decision trees and can be used for several machine learning tasks such as classification, ranking and regression. It is designated to create a fast and distributed algorithm that can handle large datasets. LightGBM is distinguished by its fast-training speed and low memory usage. In several benchmarks and experiments on public datasets, it has been shown to be faster and more accurate than XGBoost [24].

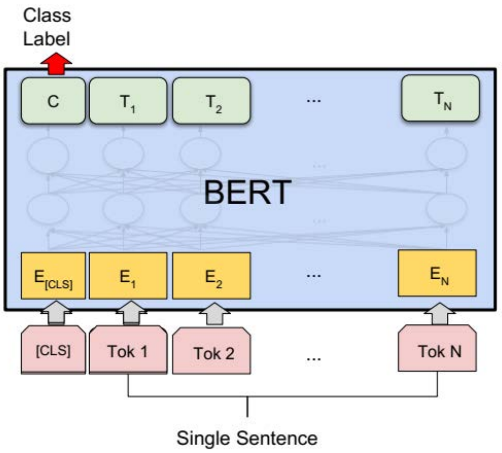

FIGURE 1. An example of single sentence classification by fine-tuning BERT [22].

LightGBM model uses leaf-wise tree growth instead of the level-wise-tree growth which is widely used in several tree-based learning algorithm such as XGBoost. In the level-based tree growth strategy, the tree structure grows level by level. In contrast, the leaf-based tree growth strategy grows the tree based on the node achieving the largest decrease in loss. In this way, the tree nodes of the leaf-wise-tree method are generally smaller than the tree nodes of the level-wise-tree method with the same tree depth, so the training process can be considerably accelerated when the dataset is large.

LightGBM uses two novel techniques which are Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). The first technique (GOSS) removes data instances with small gradients and keeps the rest to estimate the information gain. This is justified by the fact that data instances with larger gradients play a more important role in computing the information gain. Thus, GOSS can get an accurate estimation of the information gain with a much smaller data size [23]. As for the second technique (EFB), it reduces the number of features by bundling the features that are mutually exclusive [23].

The choice of LightGBM is mainly due to its great success since its appearance in 2017. This success is justified by its advantages compared to other algorithms. In fact, LightGBM is known by its fast training speed with high accuracy and low memory usage, its support of parallel, distributed and GPU learning, and its ability to handle large-scale data. Thanks to these advantages, LightGBM has become widely used in several winning solutions of machine learning competitions. In the context of malware detection, several new works have opted for this algorithm. As an example, the creators of the EMBER (Endgame Malware BEnchmark for Research) dataset [25] have chosen LightGBM as a baseline model to evaluate their dataset. They compared it with a new end-to-end deep learning model. The results showed that LightGBM outperformed the deep learning model.

C. BiLSTM

Bidirectional LSTM (BiLSTM) is an extension of the LSTM model in which training is enhanced by traversing the input data twice, left-to-right, and right-to-left. LSTM, for its part, has been proposed to solve the RNN problem which is known as ‘‘vanishing gradients’’. LSTM models extend the memory of RNN to allow it retaining and learning long-term dependencies of inputs. The LSTM memory is used to store information over a longer period of time and make decisions to preserve or ignore the information in the memory. This allows to capture important features of the inputs and preserve this information over a long period of time [26], [27].

In Figure 2, we present the architecture of BiLSTM. The contribution of BiLSTM is the application of two LSTMs on the input data. Firstly, an LSTM is applied to the input sequence in a forward direction (forward layer). Secondly, another LSTM is applied in the opposite direction of the input sequence (backward layer). The benefit from applying LSTM twice is to increase the capacity of learning long-term dependencies and subsequently improve the results of the model.

IV. THE PROPOSED MODEL

In this section we describe the intrusion and malware detection model we propose in this paper. The objective of this model is to monitor the entire healthcare environment in order to identify malware that may have penetrated the system and also detect intrusion attempts. The healthcare IT environment is heterogeneous, it is composed of several types of devices including medical and administrative staff computers and medical devices that form the so-called IoMT. With this diversity of devices, it is difficult, if not impossible, to monitor the entire healthcare environment with a single security solution. For this reason, we propose in this paper a hybrid approach to deal with the different types of attacks that this environment can encounter.

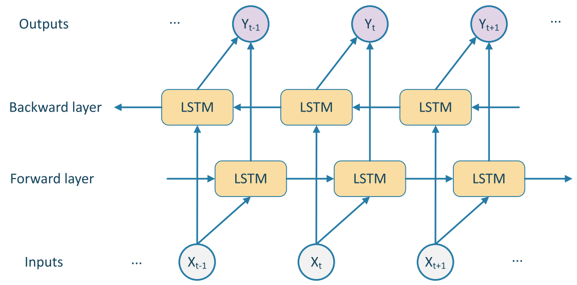

FIGURE 2. BiLSTM architecture.

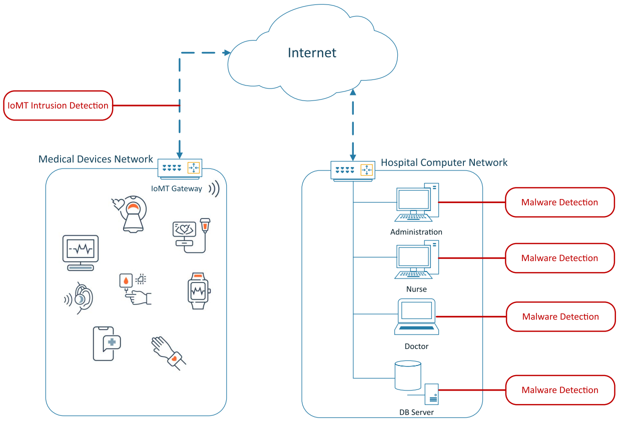

FIGURE 3. Secured healthcare network.

In Figure 3, we present the security approach that we propose for the healthcare IT environment. This approach is based on the use of two systems working together. The first one is an intrusion detection system designed to monitor the IoMT network in order to detect intrusions and malware targeting medical devices. The second is a malware detection system that should be installed on computers that are connected to the healthcare network. This module helps to monitor and analyze computer files to check the presence of malware and especially ransomware, which today represents a major challenge for the medical sector. These two systems work together to protect the entire healthcare environment despite the diversity of its components.

A. IoMT INTRUSION DETECTION

The big challenge of intrusion detection systems for medical devices is to find the right dataset to train the system. To the best of our knowledge, there are only two public datasets that are intended for this task. The first one is ECU-IoHT [13], it is a dataset designed to analyze cyber-attacks on the Internet of Health Things (IoHT). The dataset contains traces of several types of attacks that have been launched to target medical devices installed in an IoHT environment. The second dataset is ICE dataset [14], it is a network analysis of a set of ransomware attacks performed in an integrated clinical environment (ICE). This dataset is limited to a few types of ransomware that are insufficient to train an intrusion detection system. We then decided to incorporate the ECU -IoHT dataset into our model to understand IoHT attacks.

To extend the capabilities of our security model, we chose to explore two additional datasets: ToN-IoT [28] and Edge_IIoTset [29]. ToN_IoT is a novel dataset for IoT network intrusion that combines information from four heterogeneous data sources (pcap files, Bro logs, sensor data, and OS logs). The dataset testbed contains various IoT/IIoT devices and has been exposed to different types of attacks, including scanning, DoS, ransomware, backdoor, injection, XSS, password, and MITM attacks. Edge_IIoTset is a new cybersecurity dataset for IoT and IIoT applications. The data is generated from various IoT devices such as temperature and humidity sensors, heart rate sensors, flame sensors, etc. The testbed experienced 14 types of attacks including DoS/DDoS attacks, Information gathering, Man-in-the-middle attacks, Injection attacks, and Malware attacks. The two datasets are very useful for our case, as healthcare devices are part of the IoT network in general. Attacks targeting these devices are similar to attacks targeting IoT devices. To benefit from the advantages of the data offered by the three datasets mentioned above, we have opted in this approach for a hybrid solution that combines the three datasets ECU-IoHT, ToN_IoT and Edge_IIoTset. This idea allows us to address attacks on healthcare devices through the ECU -IoHT dataset and enrich the learning data with the ToN_IoT and Edge_IIoTset datasets, which are richer and contain more attacks.

Figure 4 describes the architecture of the IoMT intrusion detection system we propose in this paper. The first step consists in extracting the network flow from the captured activities, then a preprocessing phase is necessary to process the captured data, afterwards two machine learning models (Transformer-based and LightGBM) are used to classify each activity as normal or attack.

1) INPUT NETWORK FLOW

Medical devices installed in the healthcare environment are heterogeneous and have limitations related to hardware constraints. The best way to monitor the activities of these devices is to monitor the network connecting all these devices and capture each activity flowing through this network. Due to the diversity of datasets used in this approach, three types of network flows are captured for each activity. Each network flow is characterized by a set of features chosen according to the dataset description.

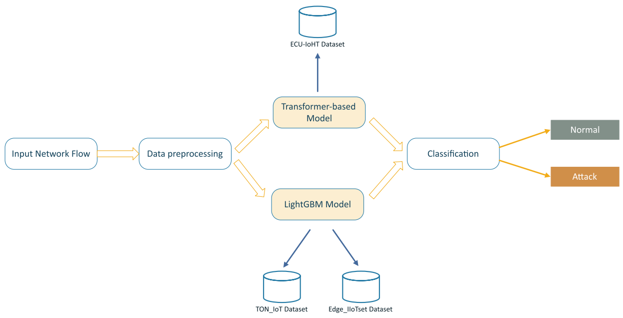

FIGURE 4. IoMT intrusion detection.

2) DATA PREPROCESSING

The preprocessing step is necessary to prepare the data for the learning phase. Since we are using two machine learning models, we need two ways to process this data.

a: DATA PREPROCESSING FOR THE TRANSFORMER-BASED MODEL

The transformer-based model is applied on textual data like the ECU-IoHT dataset. The particularity of this dataset is that it contains a textual feature describing the network activity for each input. This feature and the other features are concatenated together to form an input sequence x consisting of k words, denoted as $(x_1; x_2; \ldots; x_k)$. The transformer model used in this phase is based on the BERT pre-trained model. To pass the input sequence to this model, it must first go through two operations which are called Tokenization and Padding. The "tokenization" task consists in transforming each text into a sequence of tokens, each token must be mapped to its index in the tokenizer vocabulary of the BERT model. The "padding" operation is used to make the sequences in a uniform length.

b: DATA PREPROCESSING FOR THE LIGHTGBM MODEL

The LightGBM model is applied on datasets with numerical and categorical features like ToN_IoT and Edge_IIoTset. The objective of this phase is to eliminate unnecessary features and ensure that all data is numeric. Categorical features like protocol name, connection state, service type, dns querry, etc. are also encoded and transformed into numeric data.

3) TRANSFORMER-BASED MODEL

The objective of the transformer model is to analyze the textual data of the ECU-IoHT dataset. For this purpose, we used the BERT pre-trained model which has shown its efficiency in several NLP tasks [30]. The working principle of the transform model is described in Algorithm 1. After performing the tokenization and padding operations as described in the previous step, a special CLS token is added to the resulting sequence. The CLS token is a special component added in the BERT models to provide an aggregated representation of the whole sequence. It is typically used for classification tasks. Then the total sequence is passed as input to the BERT model. The output of this model is a sequence of hidden-states (H_S) represented as follows:

$$H_S = [C, T_1, T_2, \ldots, T_N] \tag{1}$$

where $C \in \mathbb{R}^H$ is the final hidden state vector of the special [CLS] token, $T_i \in \mathbb{R}^H$ is the final hidden state vector for the i^th input token, and H is the size of the hidden layer.

The last step in this process is to apply a fully connected neural network classifier on the final hidden state vector C of the special [CLS] token. This classifier is associated with a softmax function to map the output values to be within 0 and 1. This makes it possible to decide on the class of the input data and determine whether it is a normal or attack activity.

Algorithm 1 BERT-Based Classifier

```
Input: $x$ (text sequence)
Output: $y$ (sequence label: 0 or 1)

$x_1$ = Tokenize($x$)
$x_2$ = Pad($x_1$)
$seq$ = Add_CLS($x_2$)
last_hidden_vector = BERT($seq$)
$output$ = fully_connected(last_hidden_vector[CLS])
$y$ = softmax($output$)
```

4) LightGBM MODEL

LightGBM has shown high level performance in several machine learning tasks. It is best known for its efficiency, speed and low memory usage. These characteristics provide it a cutting edge over other machine learning algorithms. In our approach, we chose to use LightGBM to classify network flows containing numerical or categorial data such as the case of ToN_IoT and Edge_IIoTset datasets.

LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It has a number of parameters known as hyperparameters including the number of leaves per tree, the maximum depth of the tree, and the learning rate. These parameters have a considerable impact on the performance of the LightGBM model. The selection of the hyperparameters is usually done manually or through a grid searching method. In this paper we propose the integration of Bayesian optimization to search the best parameters for the lightGBM model.

Bayesian optimization is a powerful tool for optimizing objective functions that are difficult to evaluate or take a long time to evaluate [31], [32]. In our model, the optimization problem can be described as:

$$x^* = \underset{x \in \chi}{\operatorname{argmax}} f(x) \tag{2}$$

where x^* specifies the hyperparameters of the LightGBM model to be optimized. χ denotes the search space for the hyperparameters. $f(x)$ is the objective function representing the performance of the LightGBM model according to the selected hyperparameters. The evaluation criterion used to measure the performance of the objective function is accuracy. Thus, the objective of the optimization is to find the appropriate set of hyperparameters x^* allowing to obtain the maximum performance of the function $f(x)$ [33]. The optimization process usually takes place over several iterations. The objective function produces an observed result $y_i = f(x_i)$. This result is appended to the historical set $D = (x_1, y_1), \ldots, (x_i, y_i)$, which is used to update the surrogate probability model for generating the next suggestion. The surrogate probability model can be represented as follows:

$$p(y|x) \tag{3}$$

where y is the actual value of the objective function employing the proposed set of hyperparameters x, and $p(y|x)$ is the probability of y given x. This probability is useful to select the next set of hyperparameters. The optimization process of the LightGBM model is shown in Algorithm 2.

After finding the best hyperparameters, the LightGBM model is ready to start the classification process. Each instance of the input network flow is classified with the trained LightGBM model to determine if it is a normal or attack activity.

B. MALWARE DETECTION

The IoMT intrusion detection system is not sufficient to protect the entire healthcare IT environment because the majority of healthcare attacks target hospital computers through malware that can disrupt or deny the functioning of the hospital's IT system. In recent years, we have noticed a high increase in the number of ransomware attacks. This type of attack aims at blocking the hospital's system and demands ransom money to recover it. To defend against these attacks, we designed another detection system, this time aimed at identifying malware targeting hospital computers.

The purpose of malware detection is to monitor executable files on hospital computers for the presence of malware. This system uses an optimized LightGBM model to classify the input files. The training of the model is based on the EMBER dataset [25]. This dataset is useful for training machine learning models to statically detect malicious Windows portable executable files. It contains a collection of features extracted from 1.1M binary files distributed as malicious and benign. The architecture of the malware detection system is described in Figure 5. The first step is to extract features from the Windows portable executable. Then, a preprocessing operation is applied on the features. Afterwards, a classification process based on LightGBM starts to classify the executable as benign or malware.

1) FEATURE EXTRACTION

Feature extraction from portable executables is done through the LIEF parsing tool [34]. These features are extracted in accordance with the original paper of the dataset [25]. The obtained information includes five groups of features:

Algorithm 2 Bayesian Optimization of LightGBM

```
for i=1,2,3,... do
    Find a new hyperparameters configuration $x_{i+1}$ according to the acquisition function $\alpha(x)$, $x_{i+1} = \operatorname{argmax} \alpha(x, D_n)$
    Calculate the performance of $f$ according to the configuration $x_{i+1} : y_{i+1} = f(x_{i+1})$
    Update the historical set $D : D_{i+1} = D_i, (x_{i+1}, y_{i+1})$
    Update the surrogate probability model of the objective function
end for
```

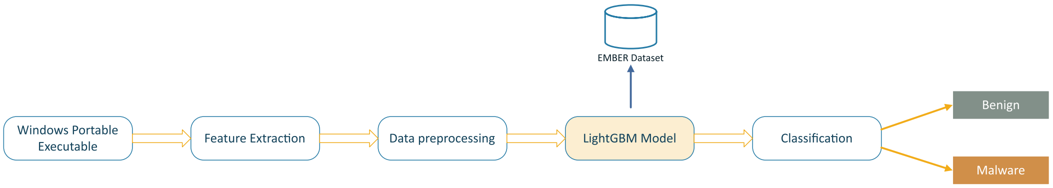

FIGURE 5. Malware detection.

- General file information: file size, virtual size of the file, number of imported and exported functions, etc.
- Header information: timestamp in the header, target machine, list of image characteristics, target subsystem, DLL characteristics, file magic, image versions, system versions, linker versions, and commit sizes.
- Imported functions
- Exported functions
- Section information: properties of some sections, such as the name, size, virtual size, and list of strings that represent the section characteristics.

Other than the mentioned features, three groups of format-agnostic information are calculated [25]:

- Byte histogram: 256 integer values representing the counts of each byte value within the file
- Byte-entropy histogram: approximates the joint distribution $p(H, X)$ of entropy H and byte value X.
- String information: contains simple statistics about printable strings included in the PE file that are at least five printable characters long.

2) DATA PREPROCESSING

The raw features extracted from the PE file are saved in a JSON file. It is therefore necessary to convert them into a model feature (vectorized features) in order to use them in the LightGBM model. This transformation is done through open-source scripts provided by the authors of the EMBER dataset. The result is a feature matrix in which each object is represented by a vector of dimension 2351.

3) LightGBM MODEL

The role of the LightGBM model is to analyze the feature vector of each PE file to decide whether it is a benign or malware file. The optimization of the hyperparameters of the LightGBM algorithm is done with the use of the Bayesian optimization technique as described previously in the LightGBM model of IoMT intrusion detection (LightGBM section).

V. EXPERIMENTAL EVALUATION

In this section, we describe the experimental evaluation of the approach proposed in this paper and present the obtained results. As described above, the proposed approach includes: (i) an intrusion detection system for IoMT based on two machine learning models LightGBM and Transformer, the first model is trained with the TON_IoT and Edge_IIoTset datasets, the second is trained with the ECU-IoHT datset; and (ii) a malware detection system based on LightGBM and trained with the EMBER dataset. To justify the choice of these algorithms, we performed a comparative evaluation consisting of testing each of the datasets with 3 machine learning techniques: LightGBM, BERT-based Transformer and BiLSTM.

A. EVALUATION MEASURES

The performance evaluation of the proposed solution is based on six standard metrics: Accuracy, Precision, Recall, F1-Score, ROC AUC and MCC. To calculate these metrics, we first need to obtain the TP, TN, FN, and FP values which are defined as follows:

- True Positives (TP): the number of attack records correctly detected.
- True Negatives (TN): the number of benign records correctly classified as normal.
- False Positives (FP): the number of normal records incorrectly classified as attack.
- False Negatives (FN): the number of attack records misclassified as normal.
- **Accuracy**: is the number of records that were correctly predicted divided by the total number of normal and attacks records.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{4}$$

- **Precision**: is the ratio of positive predictions that should be correctly classified as attack.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

- **Recall**: is the proportion of positives predictions among all positive records.

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

- **F1-Score**: is the harmonic mean of precision and recall.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

- **Receiver Operating Characteristic (ROC)** is the plot of the true positive rate (TPR) against the false positive rate FPR) at various threshold settings.
- **AUC** (Area under the ROC Curve) is an aggregate measure of performance across all possible classification thresholds.
- **MCC (Matthews correlation coefficient)** is regarded as a balanced coefficient which takes into account the four measures TP, FP, TN, and FN.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{8}$$

B. EXPERIMENTAL EVALUATION ON THE ECU-IoHT DATASET

1) DATASET DESCRIPTION

ECU-IoHT dataset is developed in an IoHT environment in order to help the healthcare security community to analyze cyberattacks on IoHT. The dataset was created through a simulated IoHT network containing some medical sensors. Different types of attacks were launched on this network. All network activities were captured and stored in the dataset as features characterizing each network flow. Table 2 shows some statistics about the dataset. It contains 23453 normal activities and 87754 attacks distributed in 4 categories: Smurf Attack, Nmap Port Scan, ARP Spoofing and DoS Attack.

2) PARAMETERS OF THE MACHINE LEARNING MODELS

Three machine learning models were used to classify the dataset. Tables 3a, 3b and 3c describe the hyperparameters used in the classification process of the three algorithms. Concerning LightGBM, the parameters were carefully chosen thanks to the Bayesian optimization technique that we included in our approach.

3) EXPERIMENTAL RESULTS

In the experimental tests of the three classifiers, we used 5-Fold Cross-Validation evaluation. The final results are calculated on the basis of the average score of the 5 folds. The results of these tests are presented in Table 4 and Figure 6. The comparative evaluation we performed showed that the three models performed well. Nevertheless, the BERT-based transformer model was the best. It obtained an ideal score of 100% in the 6 evaluation measures: Accuracy, Precision, Recall, F1-Score, ROC AUC and MCC. This can be explained by the nature of the data collected in the dataset. In fact, the network information of the activities captured by ECU-IoHT is in textual form. The transformers models are very powerful in classifying this type of data compared to other machine learning algorithms.

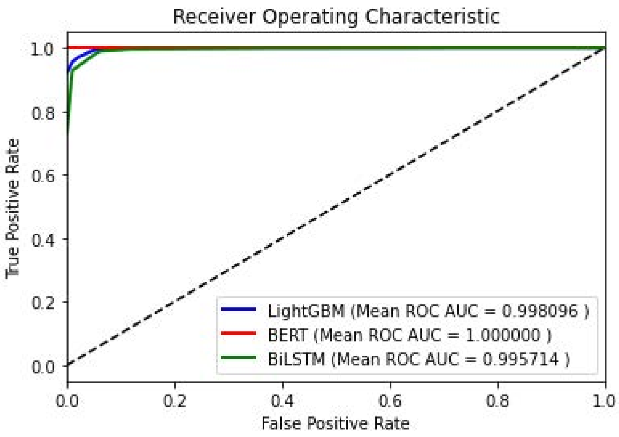

FIGURE 6. ROC curve of the 3 models on the ECU-IoHT dataset.

C. EXPERIMENTAL EVALUATION ON THE ToN-IoT DATASET

1) DATASET DESCRIPTION

ToN_IoT is a collection of IoT and Industrial IoT (IIoT) datasets designed to evaluate the effectiveness of AI-based IoT security solutions. It includes heterogeneous data sources collected from Telemetry datasets of IoT and IIoT sensors, Windows and Linux Operating systems datasets, and Network traffic datasets. Nine attack families were simulated in the datasets: Backdoor, Scanning, Ransomware, Password, Injection, DoS, DDoS, XSS, and Man-In-The-Middle attacks. In our approach we used the Train/Test Network dataset. It includes a subset of the entire attack types and normal records. It is characterized by 44 flow features extracted from the network packets, containing information about the connection, statistics, user attributes, and violation attributes. Table 5 describes some statistics regarding the concerned dataset.

2) PARAMETERS OF THE MACHINE LEARNING MODELS

The Bayesian optimization of the LightGBM model led us to choose the hyperparameters presented in Table 6. For the BERT and BiLSTM models we kept the same parameters as in Tables 3b and 3c.

3) EXPERIMENTAL RESULTS

In the experimental evaluation of the three classifiers, we used a 5-Fold Cross-Validation. We have also included the evaluation results found in the dataset article. In this article,

TABLE 2. Statistics of the ECU-IoHT dataset.

	Normal	Attacks			
Number of records	23453	87754			
		Smurf Attack	**Nmap Port Scan**	**ARP Spoofing**	**DoS Attack**
		77920	6836	2359	639

TABLE 3. Parameters of the machine learning models (ECU-IoHT)

(a) Hyperparameters of LightGBM

Hyperparameter	Value
Boosting method	gbdt
Number of iterations	411
Learning rate	0.05
Number of leaves	819
Feature fraction	0.78
Bagging fraction	0.53
Max depth	23
Min data in leaf	70

(b) Hyperparameters of BERT-based Transformer

Hyperparameter	Value
pre-trained model	bert-base-uncased
Max sequence length	64
Batch size	32
Optimizer	AdamW
Learning rate	5e-5
Number of epochs	3

(c) Parameters of BiLSTM network

Layer parameters	Value
First layer	BiLSTM with 32 units
Second layer	Dense layer with 64 units and Relu activation function
Final layer	Dense layer with 1 unit and Sigmoid activation function

TABLE 4. Experimental results on the ECU-IoHT dataset.

Classification model	Accuracy	Precision	Recall	F1-Score	ROC AUC	MCC
LightGBM	0.984182	0.985391	0.994701	0.990024	0.998096	0.952101
BERT-based Transformer	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**
BiLSTM	0.978301	0.983502	0.989094	0.986290	0.995714	0.934379

TABLE 5. Statistics of the ToN_IoT dataset.

Number of Normal records	300000	
Number of Attack records	**Backdoor**	20000
	Scanning	20000
	Ransomware	20000
	Password	20000
	Injection	20000
	DoS	20000
	DDoS	20000
	XSS	20000
	MITM	1043

TABLE 6. Hyperparameters of LightGBM (ToN_IoT).

Hyperparameter	Value
Boosting method	gbdt
Number of iterations	165
Learning rate	0.12
Number of leaves	158
Feature fraction	0.76
Bagging fraction	0.5
Max depth	19
Min data in leaf	82

the authors tested three machine learning algorithms and showed that Random Forest performed the best. All results are shown in Table 7 and Figure 7. By analyzing them we can conclude that LightGBM and BERT are the best performers with a slight lead for LightGBM. They gave results very close to 100% in the 6 evaluation measures.

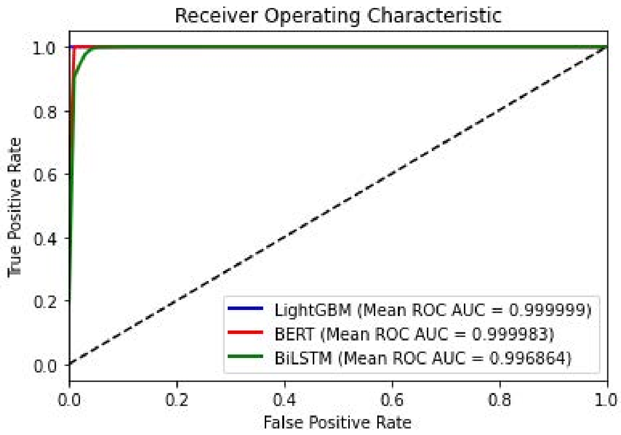

FIGURE 7. ROC curve of the 3 models on the ToN_IoT dataset.

D. EXPERIMENTAL EVALUATION ON THE Edge_IIoTset DATASET

1) DATASET DESCRIPTION

Edge_IIoTset is a new cybersecurity dataset for IoT and IIoT applications, dedicated to test intrusion detection systems. The data is generated from various IoT devices such as temperature and humidity sensors, Ultrasonic sensor, Water level detection sensor, pH Sensor Meter, Soil Moisture sensor, Heart Rate Sensor, Flame Sensor, etc. The dataset includes 14 types of attacks belonging to the following categories:

TABLE 7. Experimental results on the ToN_IoT dataset.

Classification model	Accuracy	Precision	Recall	F1-Score	ROC AUC	MCC
LightGBM	**0.999934**	**0.999888**	**0.999925**	**0.999907**	**0.999999**	**0.999856**
BERT-based Transformer	0.999852	0.999875	0.999701	0.999788	0.999983	0.999675
BiLSTM	0.973293	0.941896	0.984261	0.962611	0.996864	0.942408
Model of the dataset article (Random Forest) [28]	0.98075	-	-	0.97264	0.99688	-

TABLE 8. Statistics of the Edge_IIoTset dataset.

Number of Normal records	24301	
Number of Attack records	**Backdoor**	10195
	DDoS_HTTP	10561
	DDoS_ICMP	14090
	DDoS_TCP	10247
	DDoS_UDP	14498
	Fingerprinting	1001
	Password	9989
	MITM	1214
	Port_Scanning	10071
	Ransomware	10925
	SQL_injection	10311
	Uploading	10269
	Vulnerability_scanner	10076
	XSS	10052

TABLE 9. Hyperparameters of LightGBM (Edge_IIoTset).

Hyperparameter	Value
Boosting method	gbdt
Number of iterations	2000
Learning rate	0.05
Number of leaves	500
Feature fraction	0.8
Bagging fraction	0.8
Max depth	10
Min data in leaf	20

DoS/DDoS attacks, Information gathering, Man in the middle attacks, Injection attacks, and Malware attacks. Table 8 gives some statistics on the records of the dataset.

2) PARAMETERS OF THE MACHINE LEARNING MODELS

Like previous datasets, we have applied on Edge_IIoTset the three classification models. Concerning BERT and BiLSTM we used the same parameters as before (Table 3b and 3c). For LightGBM, the Bayesian optimization proposed us the set of hyperparameters presented in Table 9.

3) EXPERIMENTAL RESULTS

For the Edge_IIoTset dataset we have chosen to apply two classification models: binary classification (normal or attack) and multi-class classification (based on attack types). First, we start with the binary classification. The performance results of the 5-Fold Cross-Validation are summarized in Table 10. It also includes the performance results obtained from the dataset paper. The authors have evaluated the dataset with 4 different ML techniques: Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor, and Deep Neural Network (DNN). Experiments showed that DNN obtained the best accuracy. By analyzing the results mentioned in Table 10, we can conclude that LightGBM and BiLSTM obtained perfect results of value 100% in all evaluation measures, outperforming DNN which obtained an accuracy of 99.99%.

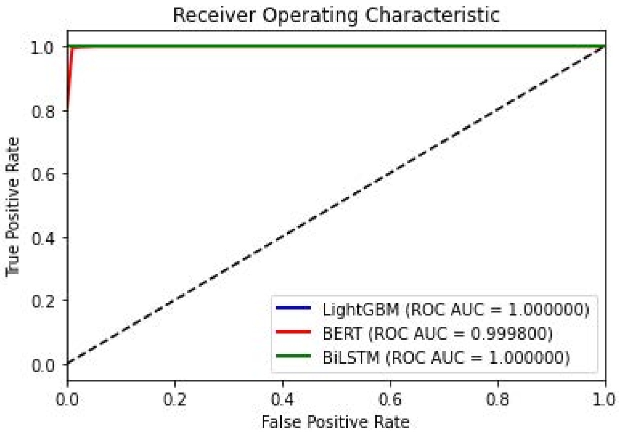

FIGURE 8. ROC curve of the 3 models on the Edge_IIoTset dataset.

To choose the best classification model for this dataset, we performed a multi-class classification to compare the 3 models LightGBM, BiLSTM, and DNN (from the dataset paper). This is a classification task based on 15 classes (normal class and 14 classes for the different types of attacks mentioned in Table 8). The results obtained are described in Table 11. Three evaluation measures are calculated for each class: precision, recall and f1- score. The results show that LightGBM performed best with an average precision of 0.92 for the 15 classes, compared to 0.90 and 0.80 for BiLSTM and DNN, respectively. LightGBM also performed best on the other metrics, with an average recall of 0.88 and an average f1- score of 0.89.

E. EXPERIMENTAL EVALUATION ON THE EMBER DATASET

1) DATASET DESCRIPTION

EMBER (Endgame Malware BEnchmark for Research) is a labeled dataset intended to train machine learning models to detect malicious Windows portable executable files statically. The dataset contains information retrieved from 1.1 million binary files, including 900,000 training samples and 200,000 test samples. The EMBER dataset contains eight groups of raw features extracted from the PE files, including five groups of parsed features (general file information, header information, imported functions, exported functions, section information) and three groups of format-agnostic information (Byte histogram, Byte-entropy histogram, String

TABLE 10. Experimental results on the Edge_IIoTset dataset.

Classification model	Accuracy	Precision	Recall	F1-Score	ROC AUC	MCC
LightGBM	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
BERT-based Transformer	0.995659	0.998411	0.996454	0.997426	0.999800	0.983672
BiLSTM	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
Model of the dataset article (DNN) [34]	0.9999	-	-	-	-	-

TABLE 11. Multi-class results for the Edge_IIoTset dataset.

	metrics	Normal	DDoS UDP	DDoS ICMP	Ransomware	DDoS HTTP	SQL injection	Uploading	DDoS_TCP	Backdoor	Vulnerability	Scanning	XSS	Password	MITM	Fingerprinting	Average
LightGBM	Precision	0.81	1.00	1.00	0.90	0.82	0.83	0.90	1.00	0.96	0.99	0.91	0.98	0.86	1.00	0.87	0.92
	Recall	1.00	1.00	1.00	0.90	0.73	0.76	0.71	1.00	0.91	0.98	0.92	0.99	0.78	1.00	0.58	0.88
	F1-score	0.89	1.00	1.00	0.90	0.77	0.80	0.80	1.00	0.94	0.98	0.91	0.98	0.82	1.00	0.70	0.89
BiLSTM	Precision	0.77	1.00	1.00	0.78	0.68	0.94	0.99	1.00	0.66	1.00	0.90	0.97	0.95	1.00	0.97	0.90
	Recall	1.00	1.00	1.00	0.48	0.86	0.68	0.67	1.00	0.86	0.97	0.89	1.00	0.72	1.00	0.57	0.84
	F1-score	0.87	1.00	1.00	0.59	0.76	0.79	0.80	1.00	0.74	0.98	0.89	0.98	0.82	1.00	0.72	0.86
DNN	Precision	1.00	1.00	1.00	0.73	0.76	0.47	0.67	0.82	0.95	0.96	1.00	0.53	0.55	1.00	0.59	0.80
	Recall	1.00	1.00	0.99	0.85	0.92	0.71	0.48	1.00	0.86	0.85	0.50	0.37	0.38	1.00	0.64	0.77
	F1-score	1.00	1.00	1.00	0.79	0.83	0.57	0.56	0.90	0.90	0.90	0.66	0.43	0.45	1.00	0.61	0.77

TABLE 12. Statistics of the EMBER dataset.

	Training	Test
Number of malicious samples	300,000	100,000
Number of benign samples	300,000	100,000
Number of unlabeled samples	300,000	–

TABLE 13. Hyperparameters of LightGBM (EMBER).

Hyperparameter	Value
Boosting method	gbdt
Number of iterations	1257
Learning rate	0.08
Number of leaves	843
Feature fraction	0.66
Bagging fraction	0.79
Max depth	23
Min data in leaf	96

information). Table 12 shows some statistics about the records in the dataset.

2) PARAMETERS OF THE MACHINE LEARNING MODELS

For the configuration of the BERT and BiLSTM models we kept the same parameters in Tables 3b and 3c, except for the Max sequence length of the BERT model, we increased it to 256 as the size of the records is larger in the EMBER dataset. Concerning the LightGBM model, the Bayesian optimization proposed us the hyperparameters presented in Table 13.

3) EXPERIMENTAL RESULTS

In the experimental evaluation we compared our optimized LightGBM model with the model proposed by the authors of the dataset. According to the source code shared by these authors [35], a LightGBM-based model with default hyperparameters was applied on the 2018 version of the EMBER dataset. The obtained ROC AUC measure is equal to 0.996428. As for our optimized LightGBM model, a ROC AUC value of 0.996830 was obtained. We can notice that our model offers a slightly higher accuracy than the baseline model. This is due to the Bayesian optimization that we have integrated in our approach. Table 14 and Figure 9 show in detail the results of the conducted experiments.

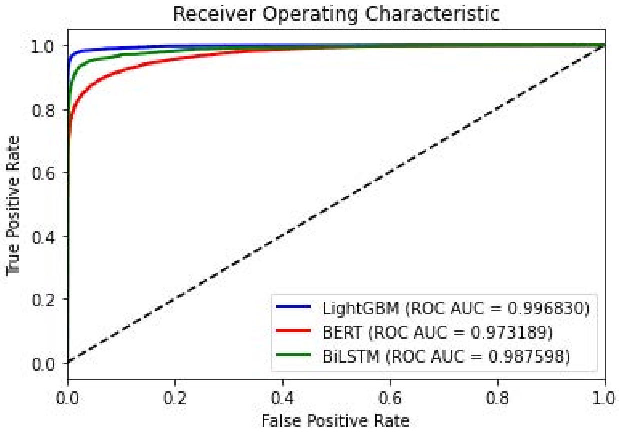

FIGURE 9. ROC curve of the 3 models on the EMBER dataset.

F. DISCUSSION

Experimental results shown that the approach we propose in this paper is very effective. The obtained results are interesting, the evaluation measures are very close to 100%. For example, the classification of the ECU-IoHT dataset showed an accuracy of 100%. The classification of ToN-IoT showed an accuracy of 99.99%. The model applied on the Edge_IIoTset dataset achieved an accuracy of 100%. Classification of the EMBER dataset using the optimized LightGBM model yielded an accuracy of 97.96% and ROC AUC of 99.68%. The experimental results also confirmed the choice of the classification models. The Transformer-based model was found to be best suited for textual data such as the ECU-IoHT dataset, outperforming the Bi-LSTM and LightGBM models. On the other hand, for categorical and numerical data (the case of ToN-IoT, Edge_IIoTset, and EMBER datasets), the Gradient Boosting models (lightGBM) are more efficient than Deep Learning models (Bi-LSTM or DNN) and the Transformer-based model in terms of accuracy and speed.

TABLE 14. Experimental results on the EMBER dataset.

Classification model	Accuracy	Precision	Recall	F1-Score	ROC AUC	MCC
LightGBM	**0.979650**	**0.983908**	**0.975250**	**0.979560**	**0.996830**	**0.959337**
BERT-based Transformer	0.912843	0.935814	0.903302	0.919271	0.973189	0.825247
BiLSTM	0.949900	0.945031	0.955370	0.950173	0.987598	0.899854
Model of the dataset article (LightGBM) [35]	–	–	–	–	0.996428	–

The strength of the proposed approach is that it has two advantages over existing work. The first is the hybrid solution that combines two detection systems to identify intrusions in IoMT networks as well as malware in Windows environments. The second advantage is the use of four different datasets to gain a better knowledge of the different types of attacks that can occur in a healthcare environment. Nevertheless, there may be a small drawback in implementing the proposed solution, which is the complexity of deployment. It requires the installation of multiple systems in different locations of the healthcare network. Moreover, a correlation operation is required to gather the obtained results. In our opinion, this limitation can be considered negligible given the good performance of the proposed solution.

VI. CONCLUSION

In this paper, a security solution for healthcare environments was presented. The solution combines two attack detection systems: (i) an intrusion detection system for IoMT networks, and (ii) a malware detection system for Windows environments. The advantage of the proposed solution is the ability to detect attacks regardless of the type of target device in the healthcare environment. Two machine learning models were used in the proposed approach: an optimized LightGBM model and a BERT-based Transformer model. These models were trained on four different datasets to ensure better knowledge of attacks on IoT, IoMT, and Windows environments. The experimental results showed that our security solution achieved a ROC _AUC score of over 99% when classifying the four datasets. In the future, we plan to extend the knowledge of our solution by adding new malware families with more complex behaviors. We also plan to design an analysis method to combine and intelligently analyze the results of our intrusion and malware detection systems. This will help to simplify the use of our solution and make accurate and fast decisions regarding the activities observed in the healthcare environment.

ACKNOWLEDGEMENT

The authors extend their appreciation to the Deanship of Scientific Research at Jouf University for funding this work through research grant no (DSR-2021-02-0108).

REFERENCES

[1] C. M. Williams, R. Chaturvedi, and K. Chakravarthy, "Cybersecurity risks in a pandemic," *J. Med. Internet Res.*, vol. 22, no. 9, Sep. 2020, Art. no. e23692.

[2] WHO. (2020). *Who reports fivefold increase in Cyber Attacks, Urges Vigilance*. Accessed: Jan. 15, 2022. [Online]. Available: https://www.who.int/news/item/23-04-2020-who-reports-fivefold-increase-in-cyber-attacks-urges-vigilance

[3] NAO. (2018). *Investigation: Wannacry Cyber Attack and the NHS*. Accessed: Jan. 15, 2022. [Online]. Available: https://www.nao.org.U.K./wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf

[4] M. Eddy and N. Perlroth. (2020). *Cyber Attack Suspected in German Woman's Death*. Accessed: Jan. 15, 2022. [Online]. Available: https://www.nytimes.com/2020/09/18/world/europe/cyber-attack-germany-ransomeware-death.html

[5] S. A. Chaudhry, A. Irshad, J. Nebhen, A. K. Bashir, N. Moustafa, Y. D. Al-Otaibi, and Y. B. Zikria, "An anonymous device to device access control based on secure certificate for internet of medical things systems," *Sustain. Cities Soc.*, vol. 75, Dec. 2021, Art. no. 103322.

[6] L. Bu, M. G. Karpovsky, and M. A. Kinsy, "Bulwark: Securing implantable medical devices communication channels," *Comput. Secur.*, vol. 86, pp. 498–511, Sep. 2019.

[7] M. Aledhari, A. Marhoon, A. Hamad, and F. Saeed, "A new cryptography algorithm to protect cloud-based healthcare services," in *Proc. IEEE/ACM Int. Conf. Connected Health: Appl., Syst. Eng. Technol. (CHASE)*, Jul. 2017, pp. 37–43.

[8] A. Yazdinejad, G. Srivastava, R. M. Parizi, A. Dehghantanha, K.-K.-R. Choo, and M. Aledhari, "Decentralized authentication of distributed patients in hospital networks using blockchain," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 8, pp. 2146–2156, Aug. 2020.

[9] P. Radoglou-Grammatikis, P. Sarigiannidis, G. Efstathopoulos, T. Lagkas, G. Fragulis, and A. Sarigiannidis, "A self-learning approach for detecting intrusions in healthcare systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.

[10] G. Thamilarasu, A. Odesile, and A. Hoang, "An intrusion detection system for internet of medical things," *IEEE Access*, vol. 8, pp. 181560–181576, 2020.

[11] S. Khan and A. Akhunzada, "A hybrid DL-driven intelligent SDN-enabled malware detection framework for internet of medical things (IoMT)," *Comput. Commun.*, vol. 170, pp. 209–216, Mar. 2021.

[12] A. Anand, S. Rani, D. Anand, H. M. Aljahdali, and D. Kerr, "An efficient CNN-based deep learning model to detect malware attacks (CNN-DMA) in 5G-IoT healthcare applications," *Sensors*, vol. 21, no. 19, p. 6346, Sep. 2021.

[13] M. Ahmed, S. Byreddy, A. Nutakki, L. F. Sikos, and P. Haskell-Dowland, "ECU-IoHT: A dataset for analyzing cyberattacks in Internet of Health Things," *Ad Hoc Netw.*, vol. 122, Nov. 2021, Art. no. 102621.

[14] L. Fernández Maimó, A. Huertas Celdrán, Á. Perales Gómez, F. García Clemente, J. Weimer, and I. Lee, "Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments," *Sensors*, vol. 19, no. 5, p. 1114, Mar. 2019.

[15] N. Garg, M. Wazid, A. K. Das, D. P. Singh, J. J. P. C. Rodrigues, and Y. Park, "BAKMP-IoMT: Design of blockchain enabled authenticated key management protocol for internet of medical things deployment," *IEEE Access*, vol. 8, pp. 95956–95977, 2020.

[16] A. A. Hady, A. Ghubaish, T. Salman, D. Unal, and R. Jain, "Intrusion detection system for healthcare systems using medical and network data: A comparison study," *IEEE Access*, vol. 8, pp. 106576–106584, 2020.

[17] J. Dong Lee, H. Soung Cha, S. Rathore, and J. Hyuk Park, "M-IDM: A multi-classification based intrusion detection model in healthcare IoT," *Comput., Mater. Continua*, vol. 67, no. 2, pp. 1537–1553, 2021.

[18] S. Shane Clark, B. Ransford, A. Rahmati, S. Guineau, J. Sorber, W. Xu, and K. Fu, "WattsUpDoc: Power side channels to nonintrusively discover untargeted malware on embedded medical devices," in *Proc. USENIX Workshop Health Inf. Technol. (HealthTech)*, Washington, DC, USA, Aug. 2013.

[19] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy*, 2018, pp. 108–116.

[20] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware images: Visualization and automatic classification," in *Proc. 8th Int. Symp. Vis. Cyber Secur.*, New York, NY, USA, 2011, pp. 1–7.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, N. Aidan Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 6000–6010.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[23] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 3149–3157.

[24] B. Quinto, *Next-Generation Machine Learning With Spark: Covers XGBoost, LightGBM, Spark NLP, Distributed Deep Learning With Keras, and More*. Berkeley, CA: Apress Jan. 2020.

[25] H. S. Anderson and P. Roth, "EMBER: An open dataset for training static PE malware machine learning models," 2018, *arXiv:1804.04637*.

[26] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 3285–3292.

[27] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," 2015, *arXiv:1511.08630*.

[28] M. Tim Booij, I. Chiscop, E. Meeuwissen, N. Moustafa, and T. H. Frank den Hartog, "ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 485–496, Jan. 2022.

[29] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022.

[30] A. Ghourabi, "SM-Detector: A security model based on BERT to detect SMiShing messages in mobile environments," *Concurrency Comput.: Pract. Exper.*, vol. 33, no. 24, p. e6452, 2021.

[31] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2010, *arXiv:1012.2599*.

[32] J. van Hoof and J. Vanschoren, "Hyperboost: Hyperparameter optimization by gradient boosting surrogate models," 2021, *arXiv:2101.02289*.

[33] X. Hao, Z. Zhang, Q. Xu, G. Huang, and K. Wang, "Prediction of f-CaO content in cement clinker: A novel prediction method based on LightGBM and Bayesian optimization," *Chemometric Intell. Lab. Syst.*, vol. 220, Jan. 2022, Art. no. 104461.

[34] Quarkslab. *Lief: Library to Instrument Executable Formats*. Accessed: Jan. 15, 2022. [Online]. Available: https://lief-project.github.io/

[35] EMBER. (2018). *Elastic Malware Benchmark for Empowering Researchers*. Github. Accessed: Jan. 15, 2022. [Online]. Available: https://github.com/elastic/ember

ABDALLAH GHOURABI received the bachelor's and master's degrees in computer science from the University of Sousse, Tunisia, in 2006 and 2008, respectively, and the Ph.D. degree in information and communication technologies from SUP'COM, University of Carthage, Tunisia, in 2014. He is currently an Assistant Professor with the Department of Computer Science, Jouf University, Saudi Arabia. He has authored and published several papers in leading international journals and peer-reviewed conferences in computer science. His research interests include cyber security, in particular, computer and network security, intrusion detection systems, malware detection, and machine learning techniques.