Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2023 doi:10.20944/preprints202302.0306.v1

Article

OCTOPUS: A Novel Approach for Health Data Masking and Retrieving Using Physically Unclonable Function and Machine Learning

Sagar Manilal Satra*
School of Engineering & Technology
Central Michigan University
satra1sm@cmich.edu

Pintu Kumar Sadhu
School of Engineering & Technology
Central Michigan University
sadhu1pk@cmich.edu

Venkata P. Yanambaka
Department of Mathematics & Computer Science
Texas Woman's University
vyanambaka@twu.edu

Ahmed Abdelgawad
School of Engineering & Technology
Central Michigan University
abdel1a@cmich.edu

*Correspondence

Abstract

The health equipment are used to keep track of significant health indicators, automate health interventions, and analyze health indicators. People have begun using mobile applications to track health characteristics and medical demands because all devices are linked to high-speed internet and phones. Such a combination of smart devices, the internet, and mobile applications expands the usage of remote health monitoring through the Internet of Medical Things (IoMT). The accessibility and unpredictable aspects of IoMT create massive security and confidentiality threats in IoMT systems. In this proposed paper - Octopus, Physically Unclonable Functions (PUFs) have been used to provide privacy to the healthcare device by masking the data, and machine learning (ML) technique is used to retrieve the health data back and for reducing security breaches on networks. This technique has exhibited 99.45% accuracy, which proves that this technique could be used to secure health data with masking.

Keywords Internet of Medical Things; physical unclonable functions; machine learning; security and privacy

1 Introduction

A person's health is critical for leading a peaceful and prosperous future. The World Health Organization (WHO) defines health as a condition of bodily and psychological well being free of sickness or disability [1]. The healthcare system is maintaining or enhancing one's health through sickness and isometric exercises, diagnosis, and cure. Most of the majority of traditional healthcare employs manual management and monitoring of patient demographic information, prior cases, diagnoses, prescription, invoicing, and pharmaceutical inventory upkeep, which results in human mistakes and negatively impacts patients. IoMT technology, mainly based on the Internet of Things (IoT), eliminates human errors. This assists physicians in diagnosing diseases more quickly and precisely by integrating all essential parameter monitoring equipment through a connection with a systematic approach [2]. The IoMT combines the IoT with medical equipment. All clinical gadgets will be linked to and analyzed by healthcare experts through internet in the future of IoMTs. This enables quicker and less expensive medical care as it develops. Figure 1 displays a case of an IoMT in which the individual vital signs are gathered by sensor devices and transmitted to the IoMT applications accessible online. The knowledge is subsequently passed to the medical professionals and personnel, who respond and then communicate with the required patients [3].

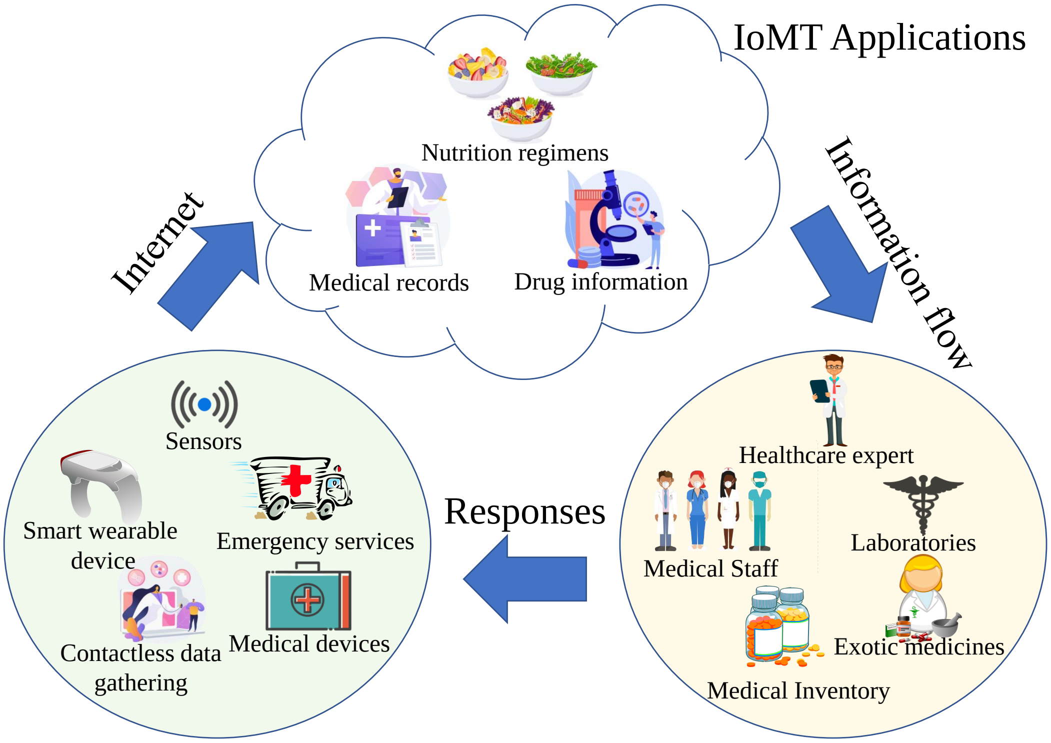


Figure 1: Internet of medical things and its applications

The IoMT is also a collection of clinical technologies and applications that connect computer networks to the healthcare profession. In the recent decade, IoMT has gotten a lot of attention. IoMT is open to attack by innumerable appearing cyber-attacks, identity fraud, keylogging, phishing, and harmful bots, as well as information processing and security issues, and data sharing across networks [4]. In 2015, an analysis by HP Fortify, the ten leading smartwatches were all discovered to have security flaws. This due to inadequate identity or permission, absence of transmitting data encryption, unsecured interfaces, vulnerable operating system, and privacy problems [5]. Validation, for instance, is confirming a user's information. All approved and verified individuals or equipment must be able to acquire IoMT devices or wireless medical devices (WMD). IoMT tools can be used for destruction in addition to abusing confidential information. For example, in a 2012 scene of the TV show Homeland, a hijacked pacemaker caused a cardiac arrest [6]. Inadequate identification safety might enable an attacker to get into the network and obtain access to a user's personal health information. Medical devices with Wi-Fi networks enable machine-to-machine communication because it is at the heart of IoMT. Numerous methods are used to minimize the total cost of managing or preventing various serious illnesses. Healthcare Information solutions have progressed, allowing consumers to keep track of patients using networks.

The importance of individual and device identification to a network is that its guarantees that data is accurately ascribed. Moreover, the data in networks is only available to authorized personnel. Devices that continually monitor healthcare attributes, devices that automatically handle therapies, and devices that track real-time data while a patient self-manages a treatment are among them [7]. Another problem with WMD is that they frequently depend on a private method for device and transmission link security. The IoMT smart devices are also a collection of sensors and electronic circuits. This helps to get the biomedical signals from the patient, network connectivity to send biomedical data via a network, and a base-band processor to analyze biomedical signals. Also, a short or long-term storage unit and a display platform with artificial intelligence systems are used for making decisions based on the doctor's availability [8].

1.1 IoMT and Smart E-Healthcare

The prevalence of internet-based computing in the healthcare industry has expanded the number of items that are connected. IoMT has applications in many fields, including smart cities, smart homes, remote health monitoring, smart healthcare, and power management. Here is a more in-depth description of IoMT and smart e-healthcare.

1.1.1 IoMT and Enabling Wireless Technologies

Individuals, society, and doctors gain from IoMT in various ways. As with any technology, there are advantages and disadvantages. Wireless technology is used to transfer health care data with the use of internet. The ample storage space provided by cloud services receives the unprocessed data been acquired at these devices/sensors [9]. To obtain a further understanding of the data collected from the devices, it is further cleaned and then analyzed. This necessitates using of new programs, devices, and tools that will improve the viewing, evaluation, transmission, and administration of the data [10].

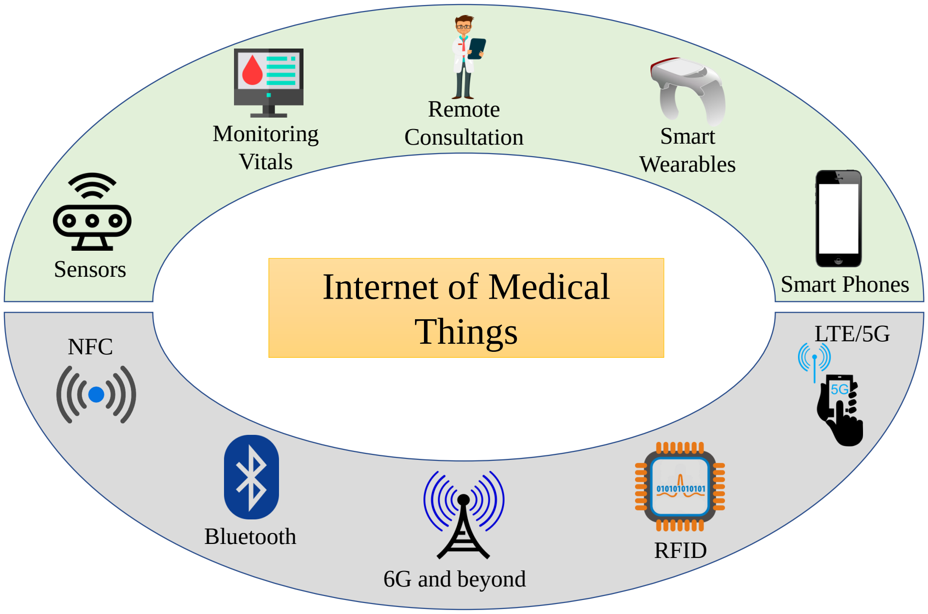


Figure 2: Internet of Medical things, enabling technologies and devices

Figure 2 depicts the interconnection of various wireless technologies, including NFC (Near Field Communications), LTE (Long Term Evolution), Bluetooth, RFID (Radio Frequency Identification), and 5G/6G (and beyond), with a variety of devices, including smartphones, multi - sensory brand, sensor systems, smart wearable, as well as other medical equipment [11]. Due to its enormous capacity and amazingly low latency advantages, 5G/6G and higher are currently widely used in IoMT.

1.1.2 Smart E-healthcare

Hospitals that use automatic and efficient modules (perhaps based on Artificial Intelligence/Machine Learning) on the Information and Communications Technology (ICT) infrastructure to enhance patient care processes and provide unique innovations are referred to as ”smart hospitals.” Intelligent hospitals have a variety of uses, including telemedicine, telehealth, and remote robotic operation [12]. Although telehealth focuses on providing non-clinical care remotely, telemedicine is used to deliver clinical treatment remotely. In remote robotic operations, medical robots carry out procedures under the direction of a surgeon who is located a great distance away.

1.2 Security and Privacy Requirements for IoMT devices

IoMT equipment has stricter security and privacy requirements than ordinary IoT-based platforms. Figure 3 shows that the patient’s data should be handled privately [13]. If the data which is collected is breached, the person can be harassed, which can also lead the patient to be distressed and depressed. Much enhanced security is needed for IoMT medical systems, such as equipment positioning, which can help to secure the system’s security and privacy. Only authorized personnel should have access to health care information, which must be collected and stored in accordance with legal and ethical confidentiality. Appropriate steps must be taken to safeguard the integrity of health information linked with patient characteristics in order to avoid data intrusions. The necessity of such safeguards cannot be overstated since data obtained by cyber-criminals might be traded on black markets, putting patients at high risk of not just privacy breaches but also financial harm [14].

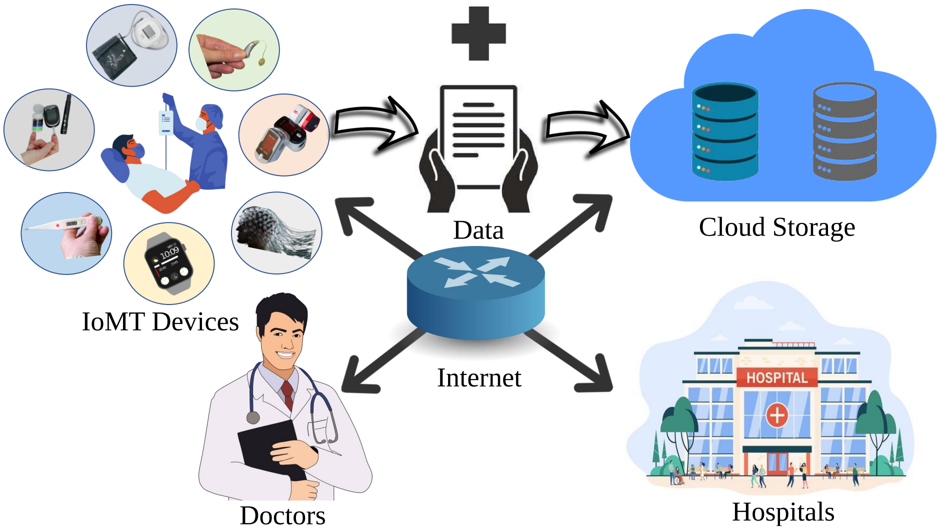


Figure 3: Security and Privacy of IoMT devices

The goal of the data authenticity condition for IoMT healthcare systems is to verify that the information received at the desired target has still not been tampered with in any manner during the information transfer. Using the broadcasting feature of the wireless connection, intruders might get information and manipulate patient records, which could have severe consequences in hazardous situations [15]. To ensure that all information has still not been tampered with, the ability to identify any unlawful information alterations or modifications is essential. As a result, proper information safety must be developed to prevent unwanted attempts to alter sent information. Furthermore, the integrity of data maintained on medical devices must be guaranteed, which implies the information cannot be tampered with [16]. WMD, particularly environmental sensors, can sometimes be seized, exposing sensitive data to hackers. Additionally, hackers can reconfigure the acquired equipment and re-deploy them to the network, monitoring conversations without really being detected [17]. As a result, medical equipment theft is a severe privacy problem that must be addressed and resolved in IoMT healthcare systems. Medical equipment in the networks must at minimal contain highly secured integrated circuits, which prohibit outsiders from reading codes placed on the equipment once they have been installed. Using PUFs to safeguard data collected in the Integrated Circuits (ICs) of medical equipment is one instance of an approach [18].

2 Related Research

Many innovative strategies for securing IoMT device data have been introduced in research articles. In recent studies, scientists claimed that cardiovascular heart disorders will kill around 23.6 million individuals by 2030. Lbrini et al. used ML to predict the cardiovascular risk present in the patient to alert them to doctors at an early stage. If a danger emerges, the doctor will get notified, and they will offer diagnosis and guidance to the patients to avoid the health consequences [19]. This study will help the patients to get the treatment as soon as possible, and also it may save lives. But this study did not cover the security part. The data can be manipulated and tampered with by an attacker to send out false information to the doctor, which can be risky. In another study, Zhao et al. analyzed hemiplegic gait based on wearable sensors. In the research, they used sensors which were worn on the patient's waist and lower limbs to collect the data while they walked in a straight line. The data was processed and analyzed to claim that the methodology they used could be used to reconstruct the patient's walking ability [20]. The security breach of the data can be avoided in both studies by masking the data. One of the studies stated that cryptographic techniques must be implemented to secure communication systems among IoMT devices. There seem to be various designs and methods that can provide privacy against a variety of integrity flaws and risks, preventing illegal access to equipment. Symmetric and asymmetric methods can be used to limit access to the device and prevent an intruder from taking control of the system. These methods can help protect against spying assaults [21]. However, if the patient's information is encrypted, it's possible that it will become a hurdle in a crisis. Whenever a healthcare professional needs to view device information, cryptographic solutions may limit access, putting the patient's life in danger.

Alternatively, fair access codes can be given to certified medical professionals managing crises [22]. However, this contradicts the purpose of having cryptographic procedures in place since attackers may acquire access to the code and manipulate it with the IoMT device via a variety of means. These can protect from external attacks, but they are sensitive to strikes from short range. Also, by flashing ultraviolet light on the individual, an attacker can acquire access to the code. The code is extracted from the gadget using electrocardiography data generated by the individual. Body interaction with the individual, on the other hand, is all that is required to obtain access to the key and remove it. Specific varieties of assaults, including radio intrusions and impersonation strikes, are also sensitive to many home automation equipment [23]. Another problem with IoMT devices is that they frequently depend on a private interface for platform and transmission channel security. No other encryption techniques are used with authentication methods, leaving the communication routes between the sensors and the controllers exposed, as stated in the preceding section. This research demonstrates a data masking mechanism that can protect against these assaults.

The proposed model Octopus approach employs PUFs, which create the cryptographic keys required for signal verification. Sensitive data components are replaced with an illegible value using masking. Since it is not a true encryption method, it is impossible to recover the original value from the disguised value. It employs a technique known as de-identification, which involves removing or masking personal identifiers like name and social security number as well as omitting or summarizing quasi-identifiers like date of birth and zip codes, or in this instance, any information pertaining to health care. Data masking is therefore one of the most often used methods for live data anonymization [24]. PUFs allow intruders to be blocked to a great extent and system protection to be reconfigured if required. PUF-based verification systems of various types have been suggested for implementation in the IoT context. In the Internet of Things, device identification is most serious concern. There have been several cases of assaults on IoT networks using rogue devices in the network undermining protection [25]. The lower cost of safeguarding a large data deployment is a vital advantage of this strategy. Masking lessens the requirement for implementing extra security controls on that data while it is stored in the platform when safe data is transferred from a secured origin into the system [26]. There are several advantages in healthcare data protection using data masking. For example - it has effective usage in clinical and healthcare industry trials, which can be useful to analyze the trials are useful for future processing. Another major advantage is that it can prevent death by protecting and securing healthcare data from attackers who can inject erroneous data. Also, data masking can help doctors to make sudden and wise clinical decisions in emergency conditions when the attacker is trying to breach the data to cause a critical situation. The main contribution of this experiment is as follows -

- PUF is used to mask health data. The classified data will be used to construct a response which is the masked data.
- Created a lightweight framework such that simple and minimal processing is required.
- Generally, after the data collection, the data is processed using mathematical operations and encryption. There is a probability of attacks exploiting the data before even the data is processed. But using the proposed methodology, the data is masked using PUF before any kind of data processing is initialized. This avoids attacks in the data collection stage.
- A simplified ML model is used to retrieve the original healthcare data with substantially less computational time, which makes this technique more suitable.
- Generally, a large dataset is used when the PUF is involved. But in the proposed experiment, the ML model is replacing the requirement of a large CRP dataset.
- The healthcare data is not sent in plaintext to the server, which makes it more secure and lightweight.
- In this proposed research, timestamps are used as a part of the challenges of the PUF. Even if the health data is identical for a particular person at different timestamps, the masked data will be different for different timestamps.
- The proposed framework gives high accuracy, which shows that it can be used to retrieve the original health data.

3 Physically Unclonable Function (PUF)

The requirement to communicate data securely has always existed. But, due to the fast expansion of digital communications, this demand has expanded drastically over time [27]. Previously, all cryptographic elements were evaluated statistically, as if they were black boxes, with prospective attackers only seeing the input and output but not the underlying activity. However, such expectations have proven exceedingly difficult to achieve in practice, necessitating loss protection measures [28]. In terms of key preservation, keys are often kept in non-volatile memory to be used in algorithms. The aim of a PUF is to produce physical elements with distinct and unexpected behavior by using unpredictable material variances created during the production process [29]. When the same challenge is presented to another PUF version, even though it was built using the same procedure, the result is different. The definition and benefits of a PUF are outlined in full below.

3.1 Definition

A PUF is concerned with the study of things that demonstrate identifiability and physical unclonability and have a challenge-response capability.

Public electronic medical equipment that is accessible by the patient is referred to as end devices. Based on the conditions, these devices are connected to a network. Information from devices will be delivered to cloud services over the internet, and the doctor will have access to it via a local area network. All the WMD has a PUF module planted in them. The production differences that arise during the creation of an integrated circuit are used in a PUF. Variations occur throughout the production process as a result of the procedures involved is included in the systems, which distinguishes them from one another. Because no two devices on a given wafer are alike, they generate various outputs. The PUF makes use of this unpredictability while generating cryptographic keys. Uncertain, uncontrolled, inevitable, and natural changes are inserted into the gadgets. As a result, the PUF device's output cryptographic keys are likewise inherently randomized. The input to a PUF is referred to as a ”Challenge,” while the output is referred to as a ”Response” [30], as shown in Figure 4.

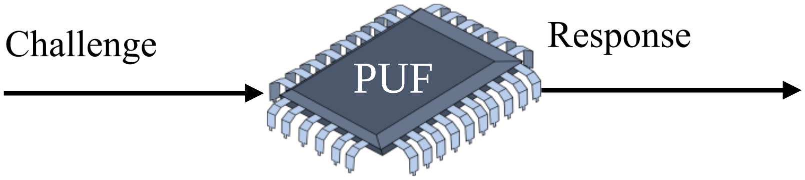


Figure 4: Challenge Response Pair

The PUF's input-output pair is known as a Challenge-Response Pair (CRP) and is often used to authenticate the gadget. For various wafers, the differences that occur throughout the manufacturing method are neither uniform nor consistent. Various PUF modules respond differently to the same challenge input. As a result, the PUF signal becomes a signature of that particular IC. The credentials are created without using any of the device's primary processor's computing capacity. As a result, the overall structure is both compact and safe. The credentials are also not kept in the sensor's memory, making them immune to numerous side-channel assaults. Based on the cryptographic protocol used or the security mechanisms used, credentials can be created as needed. This also cuts down on the amount of storage needed to enable this authentication mechanism in the environment of the IoMT. Based on the manufacturer and structure of the PUF, the number of data pairs from an individual PUF module might be extremely large. The architecture's resilience is determined by the PUF module's essential features [31], as shown in Figure 5. A controlled PUF is a PUF that has been integrated with control logic that restricts the possible evaluation methods. In general, the controlled PUF is locked without consent from a reliable source, and no answer can be effectively analyzed. More CRPs can be retrieved once a user has been granted access to one. Similar to key management, multiple session keys can be generated from a master key in this instance. The controlled PUF is constructed in practice such that the PUF and its control logic may perform complimentary functions [32], [33].

3.2 Figures of merit of a PUF

As a reliable and portable option to secure IoMT devices, PUFs are suggested by all researchers. The figures of merit of a PUF are explained below.

3.2.1 Reproducibility

Since the same challenge is presented to any PUF server, it is considered to be repeatable if the response would still be the same or nearly identical. Intra-distance is the closeness between two replies to the same challenge and PUF instance, and it is commonly assessed utilizing Hamming Distance (HD) [34].

3.2.2 Uniqueness

While the same challenge is issued to several PUF instances, their replies are significantly diverse, this is referred to as the uniqueness attribute of a PUF (i.e., their inter-distance is large). Inter-distances are frequently determined utilizing

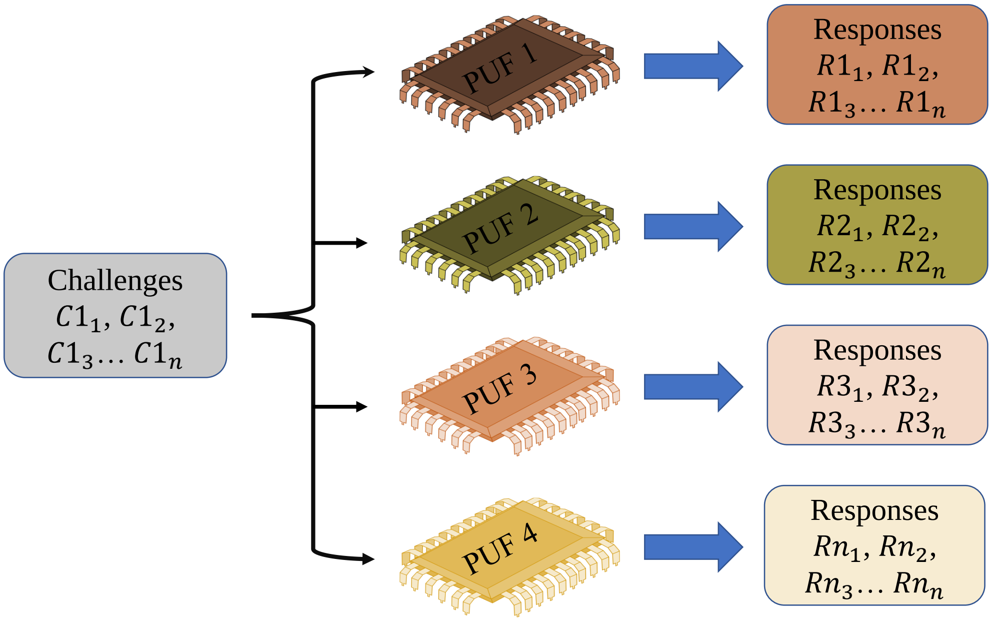

Challenges
$C1_1, C1_2, C1_3 \ldots C1_n$
PUF 1
PUF 2
PUF 3
PUF 4
Responses
$R1_1, R1_2, R1_3 \ldots R1_n$
Responses
$R2_1, R2_2, R2_3 \ldots R2_n$
Responses
$R3_1, R3_2, R3_3 \ldots R3_n$
Responses
$Rn_1, Rn_2, Rn_3 \ldots Rn_n$

Figure 5: CRP preparation

Hamming Distance in the same manner as repeatability is. Uniqueness, on the other hand, is normally assessed under standardized operating settings and not under variable situations [35].

3.2.3 Identifiability

When PUF cases can be recognized by analyzing their answers, they have the identifiability attribute. The PUF must have the traits of repeatability and uniqueness to satisfy these objectives. It's worth noting that a PUF doesn't have to be flawless in terms of repeatability or originality in order to be recognized. In reality, designing PUFs with an aggregate intra-distance of 0% is nearly difficult. Intra-distances, on the other hand, follow a confidence interval centered on a minimal value. Similarly, average inter-distances often reflect a confidence interval centered on a value somewhat less than 50%. The criteria for being recognizable is that intra-distances should be less than inter-distances with a strong chance [36].

3.2.4 Randomness

The PUF key's unpredictability is the indicator of equilibrium between the number of ones and zeros in a PUF key. A safe PUF module that can withstand extreme strength, as well as other critical guessing assaults, is one in which the PUF key contains an equal number of randomized zeros and ones [37].

4 The Proposed Model for IoMT Device

This section presents the proposed Octopus model. Figure 6 shows the elements and overview of the proposed method. Health data will be masked using the PUF on the patient's device, and data will be retrieved using a ML model in the cloud server (CS). All equipment that is introduced to the network has a PUF module built into it.

4.0.1 IoMT Device

A medical device is a resource-constrained wearable gadget that is connected to the patient. The device gathers data about the patient's condition and provides it to the doctor via the portal. In the proposed Octopus method, the WMD has a PUF module.

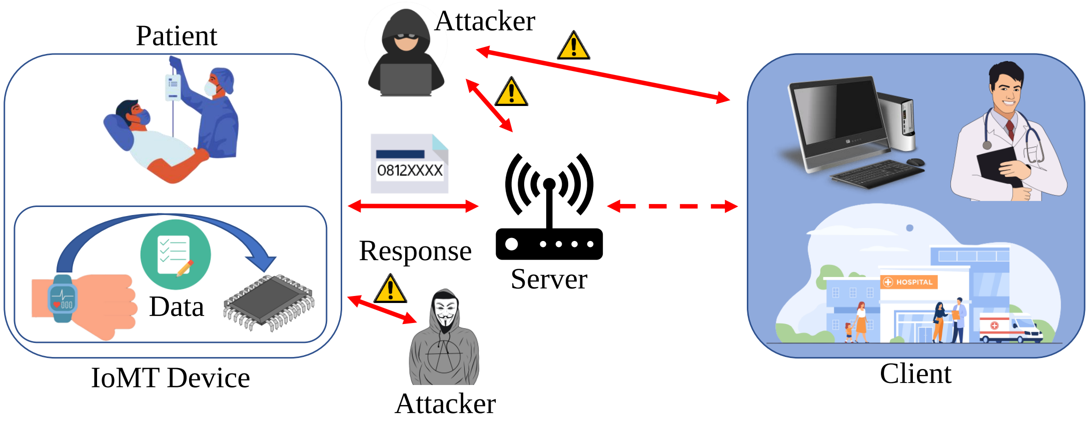


Figure 6: Overview of the Proposed Method

4.0.2 Gateway

The gateway connects the device with the CS. Data from the patient's device will be transferred to the CS using a gateway. We presume that the gateway is a reliable source of information and that it has sufficient storage and processing power.

4.0.3 Cloud Server

The CS is a centralized server to store all the patient's information. After getting masked data from the medical device, CS will retrieve the original data using ML. To retrieve a patient's health information; the doctor authenticates with the gateway using a resource-constrained smartphone or tablet. The doctor must first authenticate with the gateway in order to connect with the sensor network.

Table 1 shows the acronyms used in the method. The method is divided into two phases: 1) Enrollment phase, 2) Data masking and retrieving phase.

Table 1: Acronyms and Symbols used in the Paper

Notation	**Description**
$C, C1, C2, C3, Cn$	Challenge
$R, R1, R2, R3, Rn$	Response
$T_{st}, T_{st1}, T_{st2}, T_{st3}, T_{stn}$	Timestamp
$\rightarrow$	CRP Generation
$\ddagger$	Data Collection
$\longrightarrow$	Data transfer
$\mapsto$	ML model prediction
$\models$	Model Training
$\ni$	Database Query
$\in$	Store Operation

4.1 Assumptions

The success of data masking and retrieval was predicated on the following hypotheses.

- The WMD is incorporated with PUF chips.
- The PUFs of medical devices are strong and unaffected by outside variables like temperature, voltage, current, humidity, noise, etc.
- The ML model is only stored in the secure database (SDB) of the CS. Only the server can access the ML model to retrieve the health data.
- No CRPs will be stored anywhere.

- Before data masking, WMD is already verified in the network.

4.2 Machine Learning Algorithm

A ML classification problem with more than two classifications, or outputs, is known as multi-class classification. A multi-class classifier is a deep learning-based ML model. Since every image may be classed as many distinct animal categories as possible, using a model to identify animal species in photographs from an encyclopedia is an example of multi-class classification. Multi-class classification also necessitates the use of just one category in a dataset. Multi-class classification is perhaps the most widely used ML application, aside from regression. We are given a set of training samples separated into K distinct classes, and we create a ML method to forecast which of those classifications, some previously unknown information relates to. The model can learn characteristics unique to each class from the training dataset and utilizes those similarities to forecast the participation of the new dataset.

In the proposed framework, we have used multi-class classification which is a deep learning-based ML model. This algorithm was chosen because the healthcare data can have a range of data and each data can be leveled. Using multiclass classification, it is possible to identify the level and predict the actual health data. For example, when an IoMT device is used, it gives out a reading that can range between X to Y. This reading can be leveled as 0 to ((Y-X)-1) and this changed dataset can be trained to predict the actual IoMT device data. Multi-class classification makes the prediction that each sample is assigned to one and only one label.

The IoMT device will have a PUF in it which will collect the data from the human body and the timestamp. The data and timestamp will act as challenges of the PUF which will generate responses. The responses along with challenges (health data and timestamp) will be sent to the cloud server. The server will prepare a dataset using those data where timestamps and responses will be the input features and the IoMT device's data will be the output feature. Here ML is used for predicting and retrieving the masked health data from the IoMT device in this experiment. Multi-class classification is used in this proposed paper because healthcare data is the sequential data that is used in a particular range. The data is not in a discrete range such as housing prices, hence it can be leveled at each point of health data. Therefore, multi-class classification is a more suitable method than linear regression or any other algorithms.

4.3 Enrollment Phase

Before a device latches to a network, it is required to be enrolled on the server. The enrollment phase is divided into two steps. Figure 7 shows the process of enrolling in WMD.

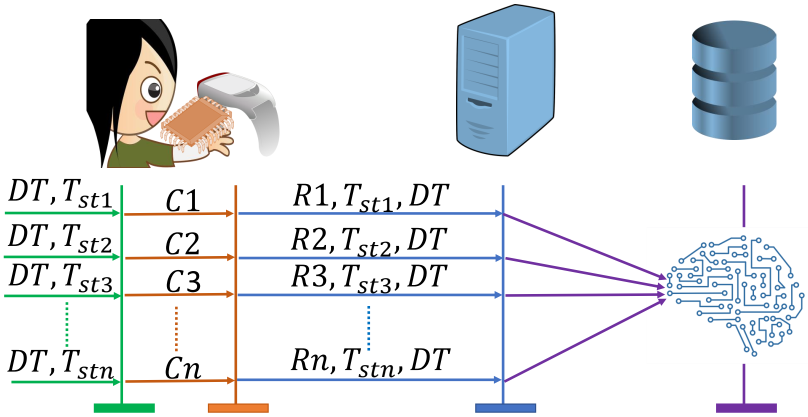


Figure 7: Enrollment Process

- **PUF response generation**: As shown in Algorithm 1, initially, DT and T_{st} will be selected, and the combination of these will act as a challenge C of the PUF of the medical device. PUF will generate a response R using the process variation of the chip. The response, timestamp, and health data will be shared with the CS through a secure communication medium.
- ML_{model} **Training and database storage**: In this step, a ML model will be trained using the received R, DT, and T_{st}. The server will use R and T_{st} as the input features and DT as the output feature. The generated model will be stored in a SDB for data retrieval. This completes the enrollment phase.

Algorithm 1: Secure Enrollment Process

```
Step-1: PUF response generation
WMD:
    $DT \,||\, T_{st} = C$
    $C \rightarrow R$
WMD $\longrightarrow$ CS $\{R, T_{st}, DT\}$
Step-2: $ML_{model}$ Training and database storage
CS:
    $R, T_{st}, DT \models ML_{model}$
CS $\longrightarrow$ SDB $\{ML_{model}\}$
SDB:
    $\in ML_{model}$
```

4.4 Data Masking and Retrieving Phase

The proposed method is presented in Figure 8. The developed scheme is divided into two steps, as shown in Algorithm 2.

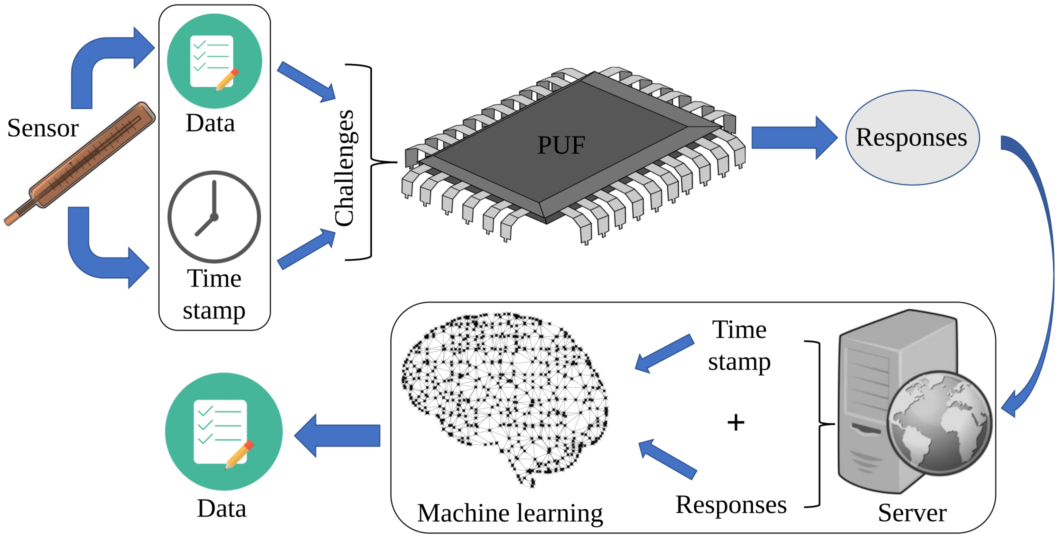


Figure 8: Proposed Model

- **Data Masking**: Data will be masked using the incorporated PUF in the WMD. At first, WMD will collect DT from the human body. Moreover, T_{st} will be identified from the clock of the WMD. Before doing any kind of further operation, both DT and T_{st} will act as challenge C of the PUF, which will generate response R. This response R will be sent to the CS using a public channel.
- **Retrieving Data**: Upon receiving R, CS will select T_{st} and will use both R and T_{st} as the input features of the stored ML_{model} for that WMD. The ML_{model} will predict the actual data DT. By following this way, CS will retrieve the masked data from the WMD.

5 Experimental results

The functionality of the suggested authentication mechanism will be examined in this section. Also, the experimental setup, dataset preparation, and ML model training and prediction will be explained in this section.

Algorithm 2: Data Masking and Retrieving Process

```
Step-1: Data Masking
WMD:
    ‡ $DT$, $T_{st}$
    $DT \,||\, T_{st} = C$
    $C \rightarrow R$
WMD $\longrightarrow$ CS $\{R\}$
Step-2: Retrieving Data
CS $\longrightarrow$ SDB $\{ID\}$
SDB:
    $ID \ni ML_{model}$
SDB $\longrightarrow$ CS $\{ML_{model}\}$
CS:
    ‡ $T_{st}$
    $R, T_{st} \mapsto DT$
CS $\longrightarrow$ SDB $\{DT\}$
SDB:
    $\in DT$
```

5.1 Experimental setup

In order to avoid the difficulties posed by the problem of latency in data processing, this project utilizes Jupyter notebook. Furthermore, the computer setup had a 3.40 GHz Intel Xeon processor, 48 GB RAM, and NVIDIA RTX A4000 32 GB GPU. The model was created and tested on a Jupyter notebook because it comes with built-in support for GPU-enabled TensorFlow and the requisite CUDA acceleration. This was done with an eye toward the model's ease of replication by the research community. There are numerous PUF architectures that can produce CRP with the necessary properties. As shown in Figure 9, the component used in this setup were an IoMT device on the client side. Raspberry pi was used as an IoMT device and the server here in this experiment. Also, a ML model is saved on the server to retrieve the data. The other component used in this research was the 64-bit arbiter PUF. The PUF was constructed using the Xilinx BASYS3 FPGA. A delay-based PUF that creates a signal based on the difference in the times of two delay lines is called an arbitrator PUF. The challenges are gathered based on a Xilinx BASYS3 FPGA's 64-bit PUF implementation. A delay-based PUF that creates a response based on the difference in the times

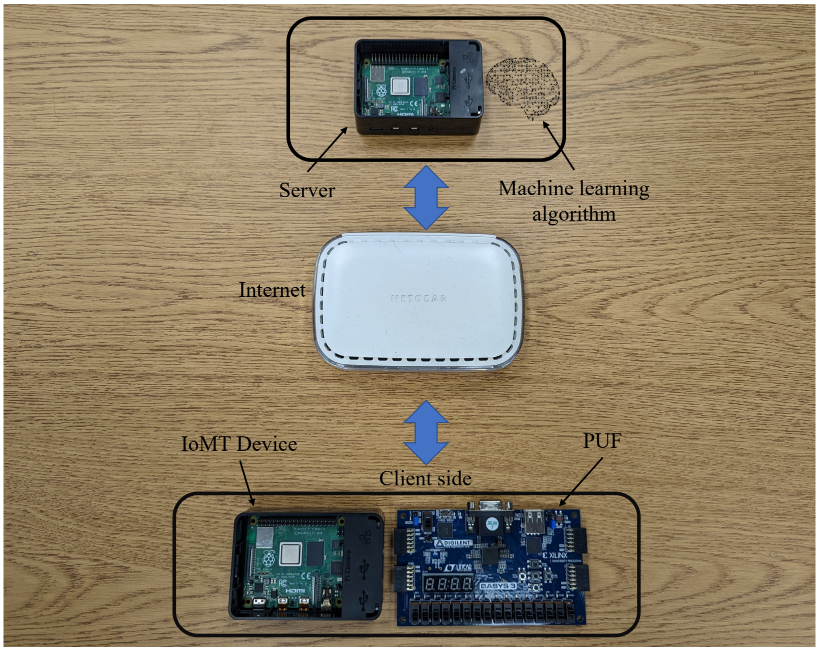


Figure 9: Experimental setup

of two delay lines is called an arbitrator PUF. The changes in the micro-electronic production process that cause a race between two identical pathways are the basis for how this PUF functions. The race has an impact on the value that the arbiter latches and is related to the variation in signal propagation latency on these two channels. The only relevant aspect of this difference is its sign, not its precise value. The sign that the arbitrator derived represents the responses and serves as the PUF identity. The arbiter can be built as a straightforward SR-latch using two cross-coupled NOR gates.

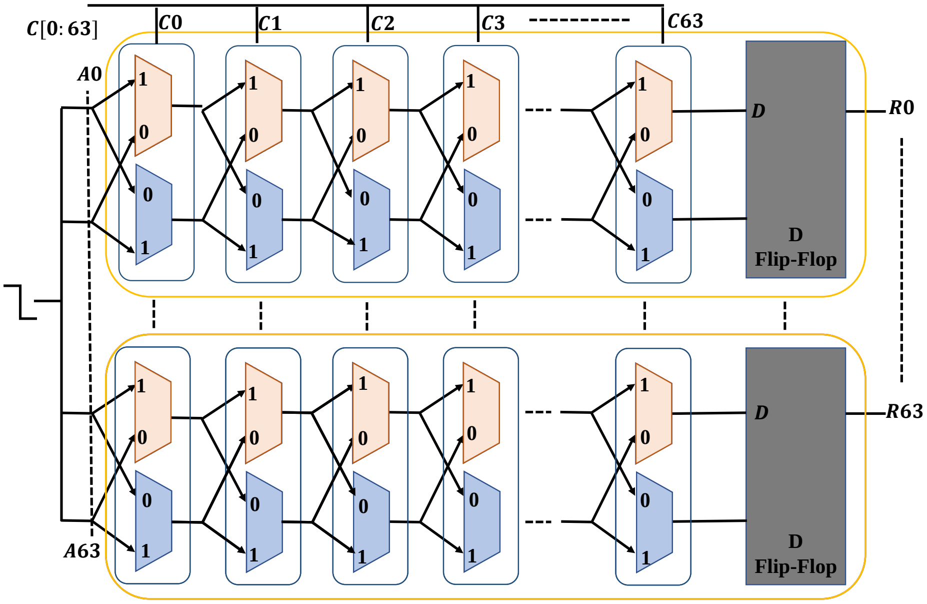


Figure 10: Arbiter PUF

Figure 10 shows that the unit of two delay lines to create a bit is represented by each box between A0 to A63. A series of red and blue colored boxes (2*1 multiplexers) is seen within each box, designating two distinct lines for processing signals. Every bit challenge will serve as a multiplexer pair's selection bit. The very first box between A0 to A63 will include challenge C0, the second box between A0 to A63 will have challenge C1, and so on. When one transmission is delivered in the PUF, the signal flows via the multiplexers in accordance with the challenge's selection bit. For instance, if C0 is 0, the transmission will flow to multiplexer C0, and if C1 is 1, the transmission will travel line 1 of the following pair of multiplexers (between red and blue). In the end, every box of the multiplexers is linked to a D flip-flop. If the flip-flop's D input receives the signal more quickly, its output indicates that the response bit would be 1. This method uses multiplexer pairs (from A0 to A63) to convert a 64-bit challenge (from C0 to C63) into a 64-bit response (from R0 to R63).

5.2 Dataset preparation

As shown in Figure 11, 64-bits was used for generating the challenges. All possible human body sample temperature was generated for an entire year. The temperature was noted at every 5 minutes interval for the whole year. The first 8-bits are used as the temperature, which is 95 to 105, and the next 8-bits are the temperature after the decimal, which are from 0.0 to 0.9. In this way, the temperature was divided into 16-bits. The next 16-bits was allocated to the months, so the 8-bits was the first digit of month which is just 0 and 1. And the next 8-bits were for the second digit of the month, which is 0.0 to 0.9. This is how the months were separated into 16-bits. The next 8-bits was for the first digit of the dates used, which is 0 to 3 and the other 8-bits was the second digit of the dates. Now for the next 8-bits, the hours in a day was used, which is 0 to 23 and for the last 8-bits, the minutes in an hour were used, which is divided by 5. So, this way, the last 8-bits were from 0 to 11. According to this, the challenges were created using python programming. In this experiment, responses were generated using a Xilinx BASYS3 FPGA's 64-bits arbiter PUF following challenges from a Raspberry PI. The challenges and the responses were merged to make a 128-bits CRP. A total of 11554460 CRPs was generated as the dataset. These CRPs were later converted into binary and are used in ML models to train the model and get the best accuracy. The first 16-bits is considered as the output feature and the rest 112-bits is considered as the input feature.

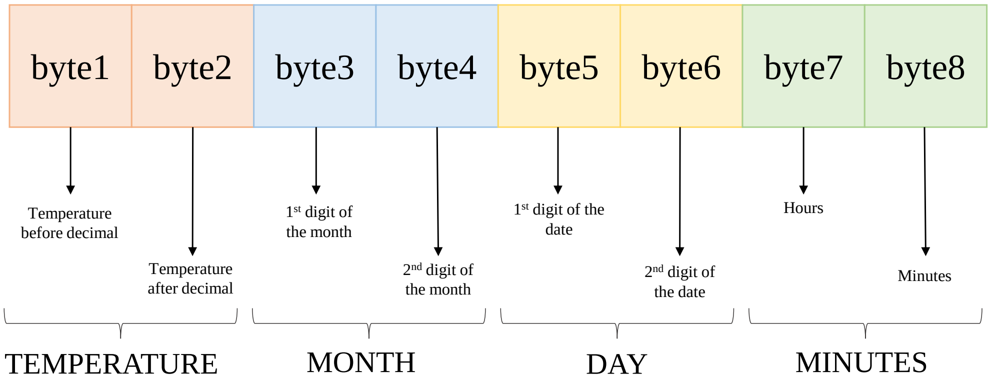


Figure 11: Data preparation format

5.3 Machine learning model training

We have used jupyter notebook to train and test the Machine learning model. In this experiment, a deep learning-based ML model (multi-class classifier) is used. A number of feeds forward, deep architectures were used for model training and evaluation. The dataset was first divided into two parts. 80% of the dataset was used to train the model, and the remaining 20% was used for validation. Categorical data was fed because the dataset had more than one discrete item, that is the temperature, which was needed as the output feature. The ML model was divided into two parts, the first part was to train the dataset for the temperature before the decimal which is 95 to 105. And the later part of the model was used to train the temperature after the decimal which is 0.0 to 0.9. Here, temperature was considered as the output feature which was 16-bit, and the rest of everything was considered as the input feature.

As shown in Table 2 many models were used for this experiment. A deep learning-based ML model which is multi-class classifier used in this experiment to retrieve the original data back. This way, the model is preserving the security and privacy of data and the user. Here, if the temperature of the patient changes, the model itself will be able to retrieve the original or changed data. In the 1st model, 4096-3072 layers were used with no dropout. But here no batch normalization was used, and the Optimizer used was ”Adam” with a total of 10 epochs and the batch size was 5000. The validation accuracy for this model was the least which is 72.06%.

Table 2: Comparison of a few models which were used for the experiment

Units	**Dropout (30%)**	**Batch normalization**	**Optimizer**	**Activation function**	**Epochs**	**Batch size**	**Validation accuracy**
4096-3072	✗	✓	Adam	Swish	10	5000	86.08
4096-3072	✗	✗	Adam	Swish	10	5000	72.06
4096-4096-4096	✗	✓	Adadelta	Swish	25	10000	94.23
4096-4096-4096	✗	✗	Adadelta	Swish	25	10000	84.73
4096-4096-4096-3072-3072-3072-2048-2048-2048	✓	✓	Nadam	Swish	50	10000	95.01
4096-4096-4096-3072-3072-3072-2048-2048-2048	✗	✗	Nadam	Relu	50	10000	88.52
4096-4096-3072-3072-2048-2048	✗	✓	RMSProp	Relu	50	5000	89.85
4096-4096-3072-3072-2048-2048	✗	✗	RMSProp	Relu	50	5000	82.08
4096-3072-3072-2048-2048-2048	✓	✗	RMSProp	Relu	50	5000	91.05
4096-3072-3072-2048-2048-2048	✓	✗	Nadam	Relu	50	5000	85.72
4096-4096-3072	✗	✓	Nadam	Relu	50	10000	89.32
4096-4096-3072	✓	✓	RMSProp	Swish	50	10000	92.05

In the 2nd model, instead of 4096-3072 layers, we used 4096-4096-4096 layers without dropouts, which was the major reason to increase the accuracy. Also, instead of using “Adam” as our optimizer we used “Adadelta” here. Unlike the 1st model we changed the epoch counts to 25 and the batch size to 10000. Because of all these changes the validation accuracy was 94.23%. For the next model, it used 4096-3072 layers without any dropout. The Optimizer utilized here was ”Adam” with a total of 10 epochs, and the batch size was 5000. Instead, batch normalization was applied and the validation accuracy here was 86.08%. The following model has a 30% dropout rate and uses 4096-4096-4096-3072-3072-3072-2048-2048-2048. The batch size was 10000, and the optimizer used was ”Nadam” with a total of 50 epochs. Batch normalization was used here to get 95.01% validation accuracy. All other models can be found in the Table 2.

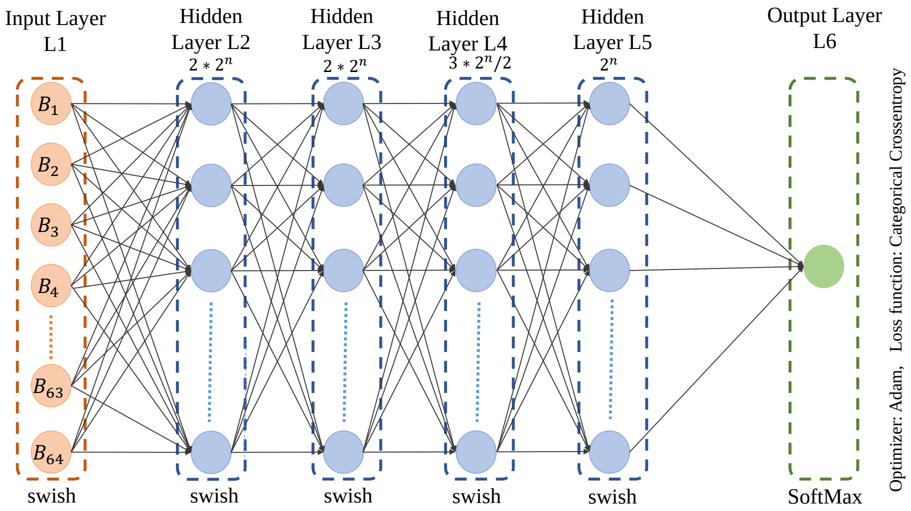


Figure 12: Multi-class classification architecture.

As the temperature series is between 95 to 105, the classes are 11. Our first ML model is comprised of four hidden layers, with $(2 * 2^n)$ neurons for the first and second hidden layer. The third hidden layer here is comprised with $(3 * 2^n/2)$ neurons. The forth hidden layer is made of (2^n) neurons as shown in the Figure 12. As the total number of class is 11, the value of n is 11.

For the model used in this experiment, 4096-4096-3072-2048 was the combination used with the layers of batch normalization. The activation function here was ”swish.” There was no dropout in this model. The optimizer used in this model was ”Adam”. ”Categorical Crossentropy” was used as the loss function because there are more than two output labels. 10 epochs in all are used, with a batch size of 5000. The metrics here is ”accuracy” as the models need to predict the output feature. As shown in Figure 13, 96.07% is the validation accuracy of this ML model, which was used for the first part to get the output feature as the temperature before the decimal.

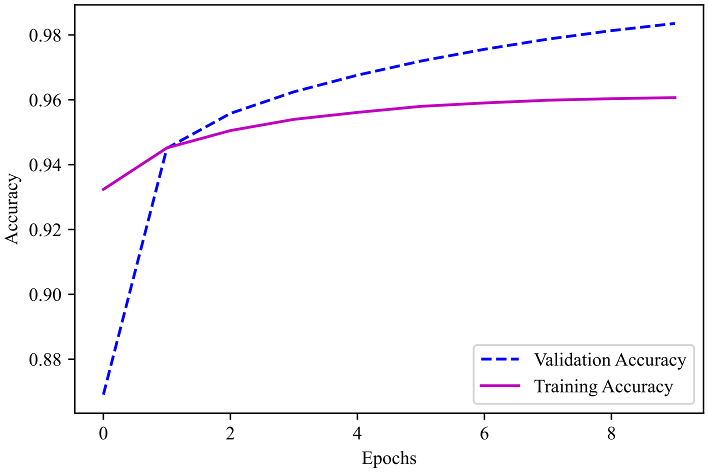


Figure 13: Validation accuracy for model 1 (Temperature 95-105)

Similarly, the second model is comprised of four hidden layers with $(4 * 2^n)$ neurons for all the four hidden layers. Also, as the decimals here were from 0 to 9, the classes are 10 and the value of n is 10. The combination utilized with the layers of batch normalization was 4096-4096-4096-4096. Here, the "swish" activation function was used. This model did not have any dropouts. "Adamax" was the optimizer employed in this model. Given that there are more than two output labels, the loss function was "categorical crossentropy." There are a total of 10 epochs used, with a batch size of 10,000. Since the models must be able to predict the output feature, "accuracy" is the key metric here. This ML model is employed for the second part's output feature, the temperature after the decimal has a validation accuracy of 89.83% as shown in the Figure 14.

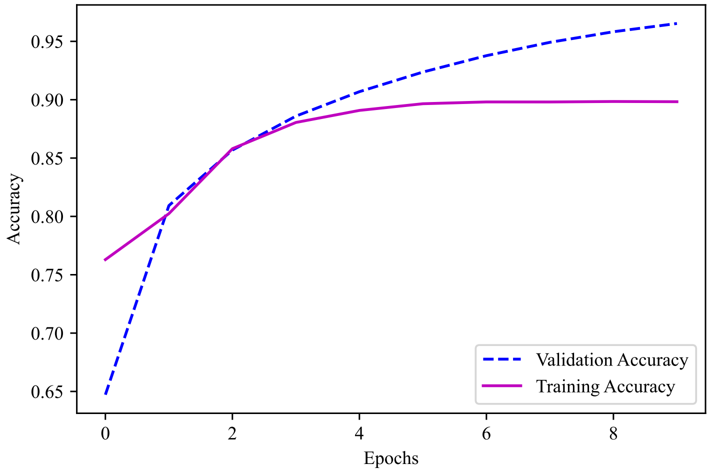


Figure 14: Validation accuracy for model 2 (Temperature 0.0-0.9)

5.4 Temperature prediction

In this experiment, the maximum accuracy found for model 1 was 96.07% and 89.83% was the maximum accuracy found for model 2. To get better accuracy different kind of method was implemented. For this part of the implementation 1,000,000 CRPs were used as the testing data. As shown in Figure 15, there were four scenarios. In all the scenarios, there were three different time stamps considered. The actual value 1 and actual value 2 are the temperatures that were selected in sequence with an interval of time. The predicted value 1 and 2 are the values that were predicted using the ML model. Also, the pre-accuracy will be counted as "1" only if the actual value is matched with the predicted values else the pre-accuracy will be counted as "0". If the pre-accuracy count is 2 or more, the accuracy will be counted as "1" and if not, the accuracy will be counted as "0".

Scenario 1

Time	Actual value 1	Predicted value 1	Actual value 2	Predicted value 2	Pre - accuracy	Accuracy
$T_{st}1$	97	98	3	2	0	0
$T_{st}2$	98	99	6	5	0	
$T_{st}3$	99	95	7	8	0	

Scenario 2

Time	Actual value 1	Predicted value 1	Actual value 2	Predicted value 2	Pre - accuracy	Accuracy
$T_{st}1$	97	97	3	3	1	0
$T_{st}2$	98	99	6	5	0	
$T_{st}3$	99	95	7	8	0	

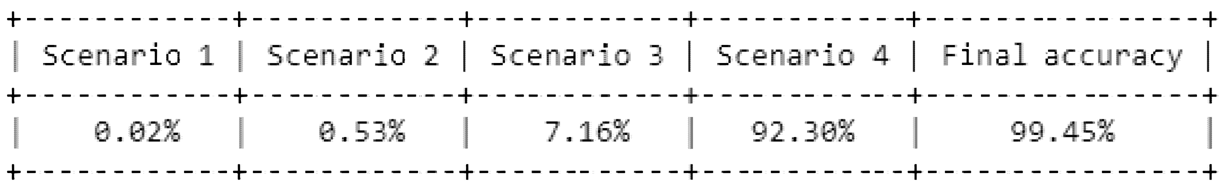
Scenario 3

Time	Actual value 1	Predicted value 1	Actual value 2	Predicted value 2	Pre - accuracy	Accuracy
$T_{st}1$	97	97	3	3	1	1
$T_{st}2$	98	98	6	6	1	
$T_{st}3$	99	95	7	8	0	

Scenario 4

Time	Actual value 1	Predicted value 1	Actual value 2	Predicted value 2	Pre - accuracy	Accuracy
$T_{st}1$	97	97	3	3	1	1
$T_{st}2$	98	98	1	1	1	
$T_{st}3$	99	99	7	7	1	

Figure 15: Temperature prediction algorithm



```
+------------+------------+------------+------------+----------------+
| Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Final accuracy |
+------------+------------+------------+------------+----------------+
|   0.02%    |   0.53%    |   7.16%    |   92.30%   |     99.45%     |
+------------+------------+------------+------------+----------------+
```

Figure 16: Experimental final accuracy

In the first scenario, none of the actual values matched the predicted values. Hence, the pre-accuracy was counted as ”0” and the accuracy, therefore was ”0”. In the second scenario, the very first actual value and the first predicted value were in agreement, but the remaining two actual values and the predicted values were not. Pre-accuracy was therefore recorded as ”1” for the first-time stamp and ”0” for the following two. Additionally, accuracy was ”0” as a result. The first and second actual values as well as the first and second predicted values were in agreement in the third scenario, however, the last actual values and the predicted values were not. As a result, pre-accuracy was noted as ”1” for the first- and second-time stamps and ”0” for the final one. As a result, accuracy was marked as ”1”. In the fourth scenario, each of the actual values and predicted values were exact matches. As a result, the accuracy was ”1” and the pre-accuracy was ”1” for each of the time stamps. A total of 1,000,000 CRPs were tested using this method and the accuracy after this implementation was found 99.45%. The result for this experiment is shown in the Figure. 16.

6 Conclusion and Future Work

Many areas of the industry have been transformed by the Internet of Things. One of the first industries to seize this opportunity by populating the internet with medical-related items is the healthcare sector. Security has emerged as the main concern as a result of rapid growth and diversity. In order to find security flaws ranging from device assaults to data transit attacks, attackers and the research community are continuously focused on the creation and rapid expansion of IoT applications. Safeguarding IoMT devices has become crucial due to the demand for using IoMT sensors to lower healthcare costs and provide better care for patients. The main issue when using networks on a wide scale is security. The confidentiality and privacy of the patients is the main area where IoMT enables smart healthcare services. In this regard, verification and permission procedures are key security requirements since they ensure that sensitive medical data cannot be intercepted. In order to overcome the constrained hardware on cloud servers, the fundamental form of distributed ML system with embedded devices was taken into consideration. To address the issue of data stealing in distributed ML systems, attempts were made to build a data-concealing framework. The experiment offered can demonstrably substantiate the validity of the data-concealing principle put out in this work. In the distributed ML system, the neural network-based model can effectively fulfill the encryption and decryption duties from the perspectives of operability and hard to steal. This is in contrast to the classic encryption technique. This paper presented a methodology that combines a mix of these strategies to achieve all security criteria since no strategy can completely satisfy the security requirements of these systems while also mitigating the majority of threats. Starting with data collection and ending with data storage and sharing, this architecture covers every stage of data and device security. Future directions and views in this area revolve around finding practical privacy and security solutions in the age of large healthcare data. In the proposed methodology, the accuracy that was found is 99.45% using ML model. In the future different federated learning models can be used to secure healthcare data. Furthermore, blockchain can be introduced to secure the healthcare. Other methods for protecting privacy must also be improved.

References

[1] Ikram Ud Din, Ahmad Almogren, Mohsen Guizani, and Mansour Zuair. A decade of internet of things: Analysis in the light of healthcare applications. *IEEE Access*, 7:89967–89979, 2019.

[2] Jerrin T John and SR Jino Ramson. Energy-aware duty cycle scheduling for efficient data collection in wireless sensor networks. *IJARCET Volume*, 2, 2013.

[3] Sahshanu Razdan and Sachin Sharma. Internet of Medical Things (IoMT): Overview, Emerging Technologies, and Case Studies. *IETE Technical Review*, 0(0):1–14, 2021.

[4] Fayez Qureshi and Sridhar Krishnan. Wearable hardware design for the internet of medical things (IoMT). *Sensors*, 18(11):3812, 2018.

[5] SD Burton, LM Tanczer, Srinidhi Vasudevan, Stephen Hailes, and Madeline Carr. The UK Code of Practice for Consumer IoT Security-Where We Are and What Next. *PETRAS, Mar*, 2021.

[6] Lavanya Sharma, Pradeep K Garg, and Sunil K Khatri. Smart E-Healthcare with Internet of Things: Current Trends, Challenges, Solutions, and Technologies. In *From visual surveillance to internet of things*, pages 215–234. Chapman and Hall/CRC, 2019.

[7] Yingnan Sun, Frank P.-W. Lo, and Benny Lo. Security and privacy for the internet of medical things enabled healthcare systems: A survey. *IEEE Access*, 7:183339–183355, 2019.

[8] Yasir Mehmood, Farhan Ahmad, Ibrar Yaqoob, Asma Adnane, Muhammad Imran, and Sghaier Guizani. Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine*, 55(9):16–24, 2017.

[9] Pintu Sadhu, Venkata Prasanth Yanambaka, Ahmed Abdelgawad, and Kumar Yelamarthi. NAHAP: PUF-Based Three Factor Authentication System for Internet of Medical Things. *IEEE Consumer Electronics Magazine*, pages 1–1, 2022.

[10] Pintu Kumar Sadhu, Venkata P. Yanambaka, Ahmed Abdelgawad, and Kumar Yelamarthi. Prospect of Internet of Medical Things: A Review on Security Requirements and Solutions. *Sensors*, 22(15), 2022.

[11] Pintu Kumar Sadhu, Venkata P Yanambaka, and Ahmed Abdelgawad. Internet of Things: Security and Solutions Survey. *Sensors*, 22(19):7433, 2022.

[12] Amir-Mohammad Rahmani, Nanda Kumar Thanigaivelan, Tuan Nguyen Gia, Jose Granados, Behailu Negash, Pasi Liljeberg, and Hannu Tenhunen. Smart e-health gateway: Bringing intelligence to internet-of-things based ubiquitous healthcare systems. In *2015 12th annual IEEE consumer communications and networking conference (CCNC)*, pages 826–834. IEEE, 2015.

[13] George Hatzivasilis, Othonas Soultatos, Sotiris Ioannidis, Christos Verikoukis, Giorgos Demetriou, and Christos Tsatsoulis. Review of Security and Privacy for the Internet of Medical Things (IoMT). In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 457–464, 2019.

[14] Fatima Alshehri and Ghulam Muhammad. A comprehensive survey of the internet of things (iot) and ai-based smart healthcare. *IEEE Access*, 9:3660–3678, 2021.

[15] Rachida Hireche, Houssem Mansouri, and Al-Sakib Khan Pathan. Security and Privacy Management in Internet of Medical Things (IoMT): A Synthesis. *Journal of Cybersecurity and Privacy*, 2(3):640–661, 2022.

[16] Bakkiam David Deebak, Fadi Al-Turjman, Moayad Aloqaily, and Omar Alfandi. An authentic-based privacy preservation protocol for smart e-healthcare systems in IoT. *IEEE Access*, 7:135632–135649, 2019.

[17] Tobias Nilges. The cryptographic strength of tamper-proof hardware. 2015.

[18] Basel Halak, Mark Zwolinski, and M Syafiq Mispan. Overview of PUF-based hardware security solutions for the Internet of Things. In *2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 1–4. IEEE, 2016.

[19] Salma Lbrini, Abdelhamid Fadil, Zakaria Aamir, Mohamed Khomali, Hassane Jarar Oulidi, and Hassan Rhinane. ”big health data: Cardiovascular disease prevention using big data and machine learning”. In *”Machine Intelligence and Data Analytics for Sustainable Future Smart Cities”*, pages 311–327. Springer, 2021.

[20] Hongyu Zhao, Haiyang Xu, Zhelong Wang, Litong Wang, Sen Qiu, Daoyong Peng, Jiaxi Li, and Jiahao Jiang. ”analysis and evaluation of hemiplegic gait based on wearable sensor network”. *”Information Fusion”*, 90:382–391, 2023.

[21] Jyoti Deogirikar and Amarsinh Vidhate. Security attacks in iot: A survey. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 32–37, 2017.

[22] Chunxiao Li, Anand Raghunathan, and Niraj K. Jha. Hijacking an insulin pump: Security attacks and defenses for a diabetes therapy system. In *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, pages 150–156, 2011.

[23] Taewan Kim, Hakjoon Lee, and Yunmo Chung. Advanced universal remote controller for home automation and security. *IEEE Transactions on Consumer Electronics*, 56(4):2537–2542, 2010.

[24] Karim Abouelmehdi, Abderrahim Beni-Hessane, and Hayat Khaloufi. Big healthcare data: preserving security and privacy. *Journal of big data*, 5(1):1–18, 2018.

[25] Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, and Jeffrey Voas. Ddos in the iot: Mirai and other botnets. *Computer*, 50(7):80–84, 2017.

[26] Lars Tebelmann, Jean-Luc Danger, and Michael Pehl. Self-secured PUF: Protecting the loop PUF by masking. In *International Workshop on Constructive Side-Channel Analysis and Secure Design*, pages 293–314. Springer, 2020.

[27] Randy Torrance and Dick James. The state-of-the-art in ic reverse engineering. In Christophe Clavier and Kris Gaj, editors, *Cryptographic Hardware and Embedded Systems - CHES 2009*, pages 363–381, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[28] Pintu Kumar Sadhu, Venkata P. Yanambaka, Saraju P. Mohanty, and Elias Kougianos. Easy-sec: Puf-based rapid and robust authentication framework for the internet of vehicles, 2022.

[29] Daniel J. Bernstein, Yun-An Chang, Chen-Mou Cheng, Li-Ping Chou, Nadia Heninger, Tanja Lange, and Nicko van Someren. Factoring rsa keys from certified smart cards: Coppersmith in the wild. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology - ASIACRYPT 2013*, pages 341–360, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[30] Venkata P. Yanambaka, Saraju P. Mohanty, Elias Kougianos, and Deepak Puthal. Pmsec: Physical unclonable function-based robust and lightweight authentication in the internet of medical things. *IEEE Transactions on Consumer Electronics*, 65(3):388–397, 2019.

[31] Muhammad Shahzad and Munindar P. Singh. Continuous authentication and authorization for the internet of things. *IEEE Internet Computing*, 21(2):86–90, 2017.

[32] Yansong Gao, Hua Ma, Said F Al-Sarawi, Derek Abbott, and Damith C Ranasinghe. PUF-FSM: a controlled strong PUF. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(5):1104–1108, 2017.

[33] Pintu K. Sadhu and Venkata P. Yanambaka. MC- PUF: A Robust Lightweight Controlled Physical Unclonable Function for Resource Constrained Environments. In *2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pages 452–453, 2022.

[34] Charles Herder, Meng-Day Yu, Farinaz Koushanfar, and Srinivas Devadas. Physical unclonable functions and applications: A tutorial. *Proceedings of the IEEE*, 102(8):1126–1141, 2014.

[35] Pintu Kumar Sadhu, Venkata P. Yanambaka, Ahmed Abdelgawad, and Kumar Yelamarthi. Performance analysis of ring oscillator puf for robust security in smart transportation. In *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*, pages 301–302, 2021.

[36] M. Garcia-Bosque, G. Díez-Señorans, C. Sánchez-Azqueta, and S. Celma. Introduction to physically unclonable fuctions: Properties and applications. In *2020 European Conference on Circuit Theory and Design (ECCTD)*, pages 1–4, 2020.

[37] Venkata KVV Bathalapalli, Saraju P Mohanty, Elias Kougianos, Babu K Baniya, and Bibhudutta Rout. PUFchain 2.0: Hardware-Assisted Robust Blockchain for Sustainable Simultaneous Device and Data Security in Smart Healthcare. *SN Computer Science*, 3(5):1–19, 2022.