



Personal Data Protection Model in IOMT-Blockchain on Secured Bit-Count Transmutation Data Encryption Approach

Sultan Almotairi^{1,*}, Santosh Reddy Addula², Olayan Alharbi³, Zaid Alzaid⁴, Yasser M. Hausawi⁵, Jaber Almutairi⁶

¹Department of Computer Science, College of Computer and Information Sciences, Majmaah University, Majmaah, 11952, Saudi Arabia

²Department of Information Technology, University of the Cumberland, Williamsburg, KY, USA

³Department of Information Systems, College of Computer and Information Sciences, Majmaah University, Majmaah, 11952, Saudi Arabia

⁴Department of Computer Science, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, 42351, Saudi Arabia

⁵IT Programs Center, Faculty of IT Department, Institute of Public Administration, Riyadh, 11141, Saudi Arabia

⁶Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia

Emails: almotairi@mu.edu.sa; santoshaddulait@gmail.com; o.alharbi@mu.edu.sa; zsazaid@iu.edu.sa; Hawsawiy@ipa.edu.sa; jalmutairi@taibahu.edu.sa

Corresponding Author: Sultan Almotairi, almotairi@mu.edu.sa;

Abstract

The Internet of Medical Things (IoMT) has paved the way for innovative approaches to collecting and managing medical data. With the large and sensitive medical data being processed hence, the need for a strong identity and privacy become necessary. The present paper suggests a comprehensive method of PriMedGuard which aims at protection of the personal medical information. The first stage will be data collection from devices and sensors, then data cleaning to transform the data into the required format. There is also a safety system in the system that registers and authenticates authorized entities as well as ETDO (Enhanced Tasmanian Devil Optimization algorithm) is used for generating asymmetric cryptographic keys. The data is encrypted using the Secure Bit-Count Transmutation (SBCT) Data Encryption Algorithm and then put in the locations provided by the InterPlanetary File System (IPFS), a decentralized and distributed storage system. A safe smart contract on the blockchain is created so that the data retrieval is secure and MedSecEnsemble Detection is proposed as an intrusion detection technique in the IoMT network. By using this method, data will stay available while at the same time integrity, confidentiality and protection against vulnerabilities are ensured. Hence, the Internet of Medical Things ecosystem will be secured from unauthorized access and possible security threats...

Keywords: Medical data analysis; Cryptography; Encryption system; Internet of Things; Ensemble Model; Blockchain technology

1. Introduction

Among many places where safeguarding personal information is critical, the link between Artificial Intelligence (AI) and collaborative learning represents an initial step of a great frontier in protecting

personal data, specifically within the broad medical field of medical records. As the digitalization of the healthcare sector goes on and the size of medical data stored in the information systems grows to an enormous volume, different remedies should be in place to be able to deal with risks such as privacy and security.

The usage of AI and machine learning technologies for protecting medical records sparks a new page in data protection. Through HIPAA compliance with the use of AI analysis taking the lead in detecting and preventing any threat to the security of medical data and at the same time, identifying complex trends. Furthermore, the collaborative learning prospects stimulate collective intelligence, enabling the two ecosystems to be able to share information and respond dynamically to the developing risks to establish a comprehensive and agile defence system. In recent years, the healthcare industry has seen significant advancements in technology, changing from conventional paper-based health records to complex electronic records and databases. This shift has indeed improved the quality of healthcare with the help of fast and accurate data transmission from healthcare professionals to patients. Nevertheless, the rapid spread of e-records has been accompanied by a range of cyber security threats and difficulties as well. Tackling these new risks necessitates the development of effective safeguards to preserve the privacy and dignity of patients. Medical records are valuable as they can contain details about a wide range of personal information such as diagnoses, treatment plans, drug history and demographic information. Data security is given very high priority, and the organization continuously employs the latest security methods both proactively and reactively.

Traditional security measures, although to some extent effective, are frequently overtaken by the perpetual evolution of cyber threats. In this aspect, the interweaving of AI and collaborative learning is paramount. The integration of Artificial Intelligence with machine learning algorithms endows those systems with the ability to analyze large amounts of data, detect abnormalities, and identify possible security threats. What distinguishes this approach from the individualized ones is that it can provide an opportunity for the interfacing and exchange of data, information, and threat intelligence in real time. This eventually helps in the development of an interconnected defence mechanism that responds to emerging threats collectively. As the incidence of cyber-criminals and the digitalization of healthcare increases, the use of AI and collaborative learning has become a strategic response that improves the security posture of health records. Such an association doesn't not only mitigate the current risks but also enables the healthcare industry to adapt to the emerging cybersecurity risks in the future. As we look at the merging of these technologies, it easily becomes evident that this new method is not just protective; it is a strategic change that shapes the future of healthcare data security. The key contribution of the paper can be summed up into the following points:

- Application of the algorithm of ETDO for the key generation leads to the strengthening of the cryptographic strength of the keys and the increase of the confidentiality and integrity of medical data.
- The implementation of the advanced SBCT Data Encryption Algorithm not only guarantees the protection of data during transmission and storage but also provides the opportunity to perform computations utilizing the encrypted data without the need for decryption which helps to preserve the privacy of sensitive medical info.
- In addition, MedSecEnsemble monitoring aims at network activity auditing and detection of potential security threats and infiltrations. The approach is instrumental in promoting the whole system's security since unauthorized access and potential security gaps are being solved.

2. Literature Review

Al-i et al. [17] built an encrypted database with homomorphic encryption and a query process with deep learning which is secure and searchable. Such a feature would make sure that users can have safe access to the PHRs that are stored in the database. The architecture consists of the user layer, the blockchain edge layer, and the IoT-based industrial layer. Three stages make up the system model: block production, checking, authentication, and registration. The Internet of Things (IoT) network is susceptible to security issues and denial of service assaults. Blockchain technology is essential in mitigating the risk of central nodes leaving the network. Multi-layer neural networks, which symbolize the mathematical computation of learning processes, are used in DL-based IDS. By finding connections between data, this method can automatically and without human intervention minimize the complexity of network traffic.

Sundas et al. [18] introduced HealthGuard, a novel security architecture for Smart Healthcare Systems (SHSs) that recognizes dangerous user actions through machine learning. To differentiate between normal and pathological activities, HealthGuard examines the vital signs of numerous SHS-connected devices. To identify potentially harmful behaviour within a SHS, four different machine learning-based detection techniques are used. Eight distinct smart medical devices were used to train HealthGuard for twelve benign events, seven of which were typical user behaviours and five of which had to do with illnesses. The anomaly detection module trains different Machine Learning (ML) algorithms to recognize abnormal activity within the SHS using the data arrays generated by the data collector module.

Mohanty et al. [19] proposed an intelligent healthcare system with secure data processing. The system has three stages: an intelligent healthcare system for patients with brain tumours, a registration module for patients, and a diagnostic module for patients. The system uses tumour detection based on in-depth learning from brain MRI and EEG signals and the modified SHA-256 algorithm. The login module allows hospital staff to access various services, including patient registration, diagnosis, pathology and enrollment. ID of patients is created for identification, and the diagnostic process begins with EEG (electroencephalography) readings and MR. (magnetic resonance imaging). Such a strategy allows the detection of healthcare insurance services fraud and enhances the patient's care. Security issues have been taken care of by the hardware of 64-bit SHA-256, making it impossible for hackers to decode. The insurance module demands the patient and the patient's insurance ID to check the health insurance information and the data must be encrypted to the monetary interests.

Kumar et. al. [20] brought up the concept of "BDSDDT" which is a Blockchain-based approach for secure data transmission in IoT-based healthcare systems. The BDSDDT proposed utilizing the ZKP mechanisms for data transmission security and integrity. It is designed to integrate with Ethereum smart contracts in a way to mitigate the data security issues and offline IPFS to manage the data storage costs. The verified data was then used to look for intrusions in the HS network. The BDSDDT framework includes several communication entities such as IoT, Edge Servers and Authentication devices. The system has two main components: a security architecture that enables blockchain and a security architecture that enables deep learning. Blockchain technology is used to register IoT devices and provide secure data transfer, while DL-enabled security architectures transform raw data into new formats, greatly reducing the dimensions of the data set.

Almalawi et al. [21] presented an encryption approach based on the Lionized remora optimization-serpent (LRO-S) optimization to cyber-attacks and privacy breaches by encrypting sensitive data. The LRO-S approach combines hybrid metaheuristic optimization and security algorithm enhancements by creating new algorithms for security key generation. The main goal is to improve the security and vulnerability of data. The study collected COVID-sensed IoT patients from remote areas and applied an LRO-based snake encryption algorithm to ensure secure transmission of data. The asymmetric hash signature ensures that only the secret key is sent to the recipient. The asymmetric hash signature function improved system security more than the encrypted hash function. The random key generation technique selects a private key, and the LRO algorithm generates a private key and a corresponding public key.

Electronic health monitoring (E-Health) systems are important for managing healthcare monitoring, but they face security issues due to potential interference from cloud storage providers (CSPs) and setups. Ghazal et al. [22] proposed an encryption framework to address these challenges using computer intelligence methods. The IoMT infrastructure recognizes patient data and sends it to blockchain technology for resource authorization and access control. Health records of patients are maintained in the database of healthcare providers by using the 'invisible' or 'unobtrusive' feature of the private blockchain. The framework comprises three phases: the first phase of private Blockchain, the training phase, and the validation phase. The blockchain of private PBCG is responsible for the PBCG and all its interactions with the device based on which all the interactions are recorded in the blockchain. This approach intends to present a reliable electronic health monitoring system and at the same time preserve the privacy of patients' electronic health records.

Through the IoT, devices are connected and cause some problems like security, reliability, and confidentiality. The proposed algorithm is a binary string search algorithm that is based on some novel group theory (BSS) and has been developed by Ali et al. [23]. The method in question provides a stable detection of the intrusion in the IoT networks and a secure search and access to the key-based databases. The newest homomorphic (EHE) encryption methods involve the following steps: installation, launching, updating, and searching. The neural network algorithms with hybrid architectures optimize

the system performance by initiating the neural network parameters, identifying the barriers and finally, evaluating the fitness function. Besides this, this process is a way of securing security locks and policy updates. This secure patient healthcare data access protocol combines a trusted network and blockchain to overcome the problems of inefficacy and security of the present scheme for sharing digital healthcare data. The proposed security search algorithm allows users to encrypt and upload them to a distributed spreadsheet. Moreover, the hybrid neural network with access control based on attributes makes it more flexible and secure.

3. Proposed PriMedGuard Methodology

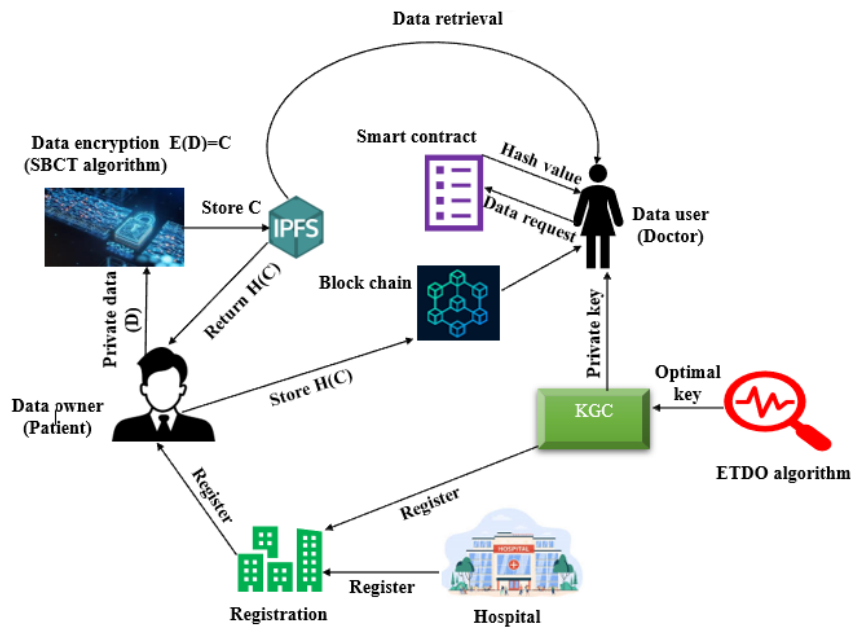


Figure 1: Architecture of the proposed PriMedGuard methodology

Comment: Can resize the text of the label used in figures. Also, add all the components used proposed model in section 3.5 of the SBT data encryption algorithm.

3.1. Data collection from the Sensing Layer:

The process begins with retrieving data from an IoMT device or sensor. These devices can be wearable health monitors, medical sensors or other connected healthcare devices that generate valuable patient data.

Comment: Details out the data collection features and process with a specific table/diagram.

3.2. Pre-processing

Ensuring data quality is very important. The previous process involved resolving missing values and normalizing data to a consistent scale. This step is the basis for preparing the data for subsequent analysis. One way to deal with missing values is to remove any rows or columns that contain null values. A column may be eliminated if more than half of its rows are null. Rows with one or more null values in their columns may also be removed. Moreover, the missing values can be handled by different mechanisms such as Deleting the entire null row and null column, replacing the null column with an arbitrary value and replacing the null values with mean, mode and median respectively.

Comment: Ensure the missing value parameters and how the system handles them. Give details of missing values handling procedures.

3.3. Registration:

A robust system for registering entities is implemented to manage access permissions. This includes the registration of authorized entities such as Patients, Hospitals, and the Key Generating Center (KGC), which is responsible for managing the cryptographic keys.

Comment: If there is a use case diagram the authors can present it can show links available on their insight.

3.4. Key Generation

Creating a cryptographic key is an important aspect of data security. Tasmanian Devil Optimization requires the use of advanced optimization algorithms for efficient and secure key generation. The optimization process aims to find optimal solutions to optimization problems similar to the Tasmanian demon nutrition process. This process involves exploring a comprehensive search area and exploiting access to optimal solutions. The Tasmanian devil's search behaviour shows the search index in the optimization process, while its pursuit process is similar to the local search exploitation.

(i) Initialization

Tasmanian devils serve as search agents in the population, forming the basis of the proposed TDO, a stochastic algorithm. A random population of these agents was first established. Based on where they are in the member search area, the TDO population is looking for a bidding site, resolving issues for candidates for problem variables. Thus, each member of the population is technically depicted as a vector. Thus, the matrix in (1) can be used to model the clusters of TDO members.

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_i \\ \vdots \\ P_M \end{bmatrix}_{M \times n} = \begin{bmatrix} p_{1,1} & \dots & p_{1,j} & \dots & p_{1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{i,1} & \dots & p_{i,j} & \dots & p_{i,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{M,1} & \dots & p_{M,j} & \dots & p_{M,n} \end{bmatrix}_{M \times n} \quad (1)$$

where P represents the population of Tasmanian devils, n signifies the number of variables, M denotes the number of searching Tasmanian devils, $p_{i,j}$ stands for its value as a candidate for the j^{th} variable, and P_i is the i^{th} potential solution. A vector is utilized to model the obtained values for the objective function as shown in Eqn. (2).

$$Fit = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_M \end{bmatrix}_{M \times 1} = \begin{bmatrix} F(P_1) \\ \vdots \\ F(P_i) \\ \vdots \\ F(P_M) \end{bmatrix}_{M \times 1} \quad (2)$$

The vector Fit represents the fitness values, with F_i denoting the value obtained by each candidate solution. Evaluating the objective function values provides insight into the quality of candidate solutions. The population's best member is determined by the candidate solution that yields the optimal objective function value. The fitness is computed using the Eqn. (3).

$$Fit = wt_1 \times C_c + wt_2 \times P + wt_3 \times K_L \quad (3)$$

where C_c , P , and K_L denotes computational complexity, primality, and key length. wt_1 , wt_2 , and wt_3 are weights that determine the significance of each criterion.

(ii) Exploration

Local carrion is a preferred food source for Tasmanian devils rather than hunting. The presence of other predators that chase after large prey and leave leftovers creates a need for alternative food sources. In these situations, Tasmanian devils choose to eat carrion. Their method of seeking carrion in their habitat can be likened to how algorithms find solutions in problem domains (key generation). The principles behind the feeding behaviour of Tasmanian devils, where they consume carcasses, are represented using mathematical models in Eqn. (4) to Eqn. (6). Each Tasmanian devil considers the locations of other members of the population as potential carrion sites within the search space. The selection of a specific location is randomly simulated in Eqn. (4), with the l^{th} population member being chosen as the targeted carrion for the i^{th} Tasmanian devil.

$$D_i = P_l, i = 1, 2, \dots, M, l \in \{1, 2, \dots, M | l \neq i\} \quad (4)$$

A new location is determined for the Tasmanian devil within the search space based on the chosen carcass. If the fitness value of the carcass is superior, then the animal moves towards it; otherwise, it

moves further away from it. This behaviour is simulated in Eqn. (5). At the final stage of this first strategy, once a new location for the Tasmanian devil has been computed, this position is established only if it yields an improved fitness value; otherwise, the animal remains at its earlier location. This process is given in Eqn. (6).

$$p_{i,j}^{T1,new} = \begin{cases} p_{i,j} + s \cdot (d_{i,j} - R \cdot p_{i,j}), & F_{Di} < F_i; \\ p_{i,j} + s \cdot (p_{i,j} - c_{i,j}), & otherwise \end{cases} \quad (5)$$

$$P_i = \begin{cases} P_i^{T1,new}, & F_i^{T1,new} < F_i; \\ P_i, & otherwise, \end{cases} \quad (6)$$

Here, $P_i^{T1,new}$ represents the updated status of the Tasmanian devil considering a new strategy, R representing a randomly generated number that can be either 1 or 2, F_{Di} denotes the selected carrion's objective function value, with s indicating a random number within the interval, $F_i^{T1,new}$ represents its newly calculated objective function value, while the value $p_{i,j}^{T1,new}$ is determined for the j^{th} variable based on the first strategy

(iii) Exploitation

Hunting and foraging are the second livelihoods of the Tasmanian Devil. There are two stages for Tasmanian demonic activity during the offensive. It finds the prey and attacks it at an early stage by scanning it around. Then, in the 2nd stage, it chases the prey to stop and starts feeding later on. Eqn. (7) to (9) are used to mimic the selection of prey and attack behaviour.

In the second strategy, the position of the prey is assumed from the position of the other members of the population while updating the Tasmanian demon. The k th member of the population is randomly selected to be wild, and l is a naturally occurring random integer between 1 and M opposite i . In Eqn. (7) the process of prey selection is simulated.

$$Q_i = P_l, i = 1, 2, \dots, M, l \in \{1, 2, \dots, M | l \neq i\} \quad (7)$$

where, Q_i is the selected prey by the i^{th} Tasmanian devil.

When the position of the prey is confirmed, the new location of the Tasmanian devil is calculated. If the intended fitness value of the chosen prey is higher, it enters this new place; otherwise, it moves out of that location. This process is depicted in Eqn. (8). If the new location for the Tasmanian demon increases the value of the target function, it replaces the old one as given in (9).

$$p_{i,j}^{T2,new} = \begin{cases} p_{i,j} + s \cdot (q_{i,j} - R \cdot p_{i,j}), & F_{Qi} < F_i; \\ p_{i,j} + s \cdot (p_{i,j} - q_{i,j}), & otherwise \end{cases} \quad (8)$$

$$P_i = \begin{cases} P_i^{T2,new}, & F_i^{T2,new} < F_i; \\ P_i, & otherwise, \end{cases} \quad (9)$$

where $P_i^{T2,new}$ represents the updated status of the i^{th} Tasmanian. F_{Qi} signifies the fitness of the chosen prey, $F_i^{T2,new}$ refers to its newly calculated objective function value, and the value for the j^{th} variable is denoted by $p_{i,j}^{T2,new}$.

A local search of a search area is similar to a wildlife search near an attack site. This Tasmanian devil's behaviour demonstrates how TDO can be used to unify top-candidate solutions. Tasmanian devils chase victims around the location of the attack to mimic the search process. The Tasmanian devil's chase phase is modelled using Eqn. (10) to (12). At this point, the position of the Tasmanian devils is considered to be the centre of the neighbourhood where wildlife exploration takes place.

$$RAD = 0.01 \left(1 - \frac{l}{I_{max}} \right) \quad (10)$$

$$p_{i,j}^{new} = p_{i,j} + (2s - 1) \cdot RAD \cdot p_{i,j} \quad (11)$$

$$P_i = \begin{cases} P_i^{new}, & F_i^{new} < F_i; \\ P_i, & otherwise, \end{cases} \quad (12)$$

where P_i^{new} denotes the updated status of the i^{th} Tasmanian devil in proximity to P_i , $p_{i,j}^{new}$ signifies its new value for variable j , and F_i^{new} indicates its newly calculated objective function value. I_{max} is the maximum number of iterations allowed, I stands for the current iteration count, and RAD represents the radius of the neighbourhood centred on a point that has been attacked.

3.5. Proposed SBCT Data Encryption Algorithm:

To protect sensitive medical data during transmission and storage, homomorphic encryption is performed. This encryption technique allows the calculation of encrypted data without the need for decryption, increasing the security and privacy of the entire data.

3.5.1. Encryption

Key Generation:

- Generate a cryptographic key (K) of any length, ensuring its randomness and security.

Text Conversion to Binary Matrix (M):

- Convert each character to its corresponding ASCII value.
- Convert the ASCII value to binary form.
- Pad with zeros to ensure a fixed length of 8 bits.
- Randomly adjust the length to 8 bits if necessary.
- Form a matrix 'M' by placing the binary values of each character in rows.

Bit Transmutation Operations:

Introduce a set of unique bit transmutation operations:

- Chaotic Shift (CS)-Randomly shuffle the bits in each row of matrix 'M'.
- Quantum Swap (QS) - Swap specific bits in the matrix 'M' based on a pseudo-random pattern.
- Dynamic Ripple (DR)-Create a ripple effect by propagating changes in bits from one row to another in a dynamic pattern.
- Bitwise Distortion (BD)- Apply a distortion function to randomly alter certain bits in the matrix 'M'.

Dynamic Rotation Phase:

Based on a unique interpretation of the key (K):

- If the first digit is 'EVEN', apply CS as the number of times as the second digit in the key.
- If the first digit is 'ODD', apply QS as the number of times as the second digit in the key.
- If the third digit is 'EVEN', apply DR as the number of times as the fourth digit in the key.
- If the third digit is 'ODD', apply BD as the number of times as the fourth digit in the key.

Matrix Transposition and Flattening:

- Transpose the resultant matrix 'M' and store it in 'MT'.
- Flatten the matrix 'MT' by reading it column-wise to create a linear array 'T'.

Bit Count Transmutation:

- Until a different bit is found, count similar bits in 'T'.
- Repeat the process until the end of 'T'.

This algorithm introduces a combination of unique bit transmutation operations, dynamic rotations, and a novel approach to bit counting for enhanced security and unpredictability. The distinctiveness of the operations contributes to the robustness of the algorithm against various cryptographic attacks.

3.5.2. Decryption

Retrieve the cryptographic key (K) used during the encryption phase.

Bit Count Inversion:

- Parse the encoded message generated during encryption, reading bit count followed by the previous bit and the unique delimiter.
- Invert the bit count to reconstruct the original bit sequence.
- Repeat the process until the end of the encoded message.

Matrix Construction from Linear Array (T):

- Reconstruct the transposed matrix 'MT' by reading the linear array 'T' column-wise.

Dynamic Rotation Inversion:

Based on the interpretation of the key (K) during encryption:

- If the third digit was 'EVEN', apply the inverse of Operation C (Dynamic Ripple) as many times as the key's fourth digit.
- If the third digit was 'ODD', apply the inverse of Operation D (Bitwise Distortion) as many times as the key's fourth digit.
- If the first digit was 'EVEN', apply the inverse of Operation A (Chaotic Shift) as many times as the key's second digit.
- If the first digit was 'ODD', apply the inverse of Operation B (Quantum Swap) as many times as the key's second digit.

Matrix Inversion-Inverse Operations:

Apply the inverse of the unique bit transmutation operations used during encryption:

Inverse Operation A (ICS)- Inverse Chaotic Shift:

Reverse the random shuffling of bits in each row.

Inverse Operation B (IQS)- Inverse Quantum Swap:

Reverse the specific bit swaps based on the pseudo-random pattern.

Inverse Operation C (IDR)-Inverse Dynamic Ripple:

Undo the ripple effect by propagating changes in bits from one row to another.

Inverse Operation D (IBD)- Inverse Bitwise Distortion:

Reverse the distortion function applied to alter certain bits.

Binary Matrix to Plain Text:

Convert the binary matrix 'M' back to its original form by converting each row to ASCII values.

Concatenate the ASCII values to retrieve the plain text.

Algorithm 1: SBCT Data Encryption & Decryption Algorithm

SBCT Data Encryption & Decryption Algorithm

```

Start ();
{
  Encryption ();
  Key generation:
  Generation of cryptographic key (K)
  Text to binary conversion ();
  Conversion of characters by ASCII value;
  ASCII→binary conversion:
  Length ();
  Adjusting the data length by 8-bits
  Provides binary values to the matrix 'M':
}
{
  Bit Transmutation ();
  Introducing unique bits for transmutation operation;
  Chaotic Shift;
  Randomly shuffle the bits
  Quantum Swap;
}

```

```

    Bit swapping based on pseudo-random pattern
Dynamic Ripple;
    Developing ripple effect over the bits
Bitwise Distortion;
    Randomly altering certain bits
Dynamic Rotation Phase ();
    If the first digit is EVEN;
        Apply CS→second digit in the key
    If the first digit is ODD;
        Apply QS→second digit in the key
    If the third digit is EVEN;
        Apply DR→ forth digit in the key
    If the third digit is ODD;
        Apply BD→ forth digit in the key
}
{
Transposition and Flattening of Matrix ();
     $M \rightarrow M^T$ 
     $M^T \rightarrow$  to linear  $T$ 
Bit count transmutation ();
    similar bit count in  $T$  until different bit found
    Repeat until found a new bit
End ();
}
Stop ();
Decryption ();
Start ();
{
    Retrieving the cryptographic key ();
Bit Count Inversion ();
    Encrypted bit ← reading the previous bit count
    Bit inversion ← reconstructing the original bit sequence
    End ← after getting an encoded message bit
Matrix construction ();
    Reconstruction of  $M^T \rightarrow T$ 
}
{
Dynamic rotation inversion ();
    If the third digit is EVEN;
        Apply →  $DR^{-1}$  as fourth digit in the key
    If the third digit is ODD;
        Apply →  $BD^{-1}$  as fourth digit in the key
    If the first digit is EVEN
        Apply →  $CS^{-1}$  as second digit in the key
    If the first digit is ODD
        Apply →  $QS^{-1}$  as second digit in the key
}
{
Matrix Inversion-Inverse Operations ();
     $CS^{-1} \rightarrow$  Random shuffling of bits
     $IQ^{-1} \rightarrow$  Bits swapping based on pseudo-random pattern
     $DR^{-1} \rightarrow$  Changing the bits from one row to another
     $BD^{-1} \rightarrow$  Altering the bits by reverse the distortion function
Text conversion ();
    Conversion of binary matrix;
    Binary matrix → original form → ASCII
    ASCII → plain text
}
end ();
Stop ();

```

3.6. IPFS - Data Storage:

Comments: data sources and samples are not properly described.

The Interplanetary File System, or IPFS for short, is a file storage system. It makes it simple to exchange and safely store different kinds of data by giving each file a hash value depending on its content. This is especially important, as it is not location-based addressing depended. In addition, IPFS makes use of deduplication techniques hence avoiding the duplication of storage thus saving space and reducing the costs and improving the speed at which the records can be accessed. Moreover, by making use of a hash value after a file has been hashed, it can no longer be altered. Tampered with. The fact that IPFS is a protocol for storage means that it will be the basis for sharing and permanent storage of files, where Electronic Health Records are also included.

3.7. Data Retrieval- Blockchain Smart Contract:

Blockchain is a decentralized and disruptive technology; smart contracts, it is used for getting secure and audited data. This guarantees that only authorized entities can have access to some data enhancing publicity and accountability overall.

3.8. Detection of intrusions using the proposed MedSecEnsemble techniques:

Machine learning algorithms are applied to discover any suspicious patterns or potential intrusions in IoMT systems. This is very vital to uphold the privacy and security of the medical data and deny unauthorized access or disturbance.

After the data preprocessing step is finished, the next important step is to use machine learning methods to detect intrusions, in the IoMT networks. This specific point is very important to protect data from unauthorized access or disruptions.

Machine learning algorithms are applied to the processed data to find out any abnormal behaviours or possible attacks inside the networks of IoMT. These practices include the training of models to identify network behaviours and flagging any deviations that might be a case of a security breach or intrusion.

3.8.1 Base Classifiers:

Comment: Can add a mathematical description of the classifiers used below

(i) Isolation Forest (IF):

The IF algorithm targets the detection of unusual data points in a data set using the innovative anomaly detection method based on the isolation principle. While training, IF constructs a library of decision trees, where each tree randomly picks features and uses them to split and isolate instances. The length of the paths between each case in the tree is calculated to get an anomaly score; a short path means a higher chance of anomaly. If technology can deal with large data, its versatility makes it possible to be applied in many areas such as network security, fraud detection, healthcare, and manufacturing industries (for quality control). Furthermore, its effectiveness in high-dimensional feature spaces has led to widespread adoption for anomaly detection tasks. However, the anomaly detection by isolation forest algorithm can be deliberated using eqn. (13),

$$X(s, n) = 2^{-\frac{F(g(X))}{d(N)}} \quad (13)$$

Where the term $d(N)$ refers to the total amount of data selected for detection, $g(X)$ defines the characteristics of standard data and (s, n) are the parameters used for observation respectively.

(ii) One-Class SVM (OCSVM)

One-Class Support Vector Machine is an ML algorithm used for anomaly detection in healthcare. It belongs to the category of unsupervised learning and operates by identifying data points that deviate significantly from the norm within a dataset. This method builds the discriminative decision boundary around the normal instances of the data set, aiming to separate them from possible outliers, consequently, an effective outlier detection can be achieved. OCSVM use the kernel function to create the higher-dimensional spaces and then the optimal hyperplanes are used to separate the normal and

abnormal observations. To attain this purpose, OCSVM can be used effectively to identify out-of-the-ordinary patterns or inconsistencies in health data sets and, therefore, it can be used in the detection of abnormalities.

(iii) LSTM Autoencoder

The LSTM Autoencoder, a type of deep learning model involving unsupervised learning for sequence data analysis, is a subset of the family of autoencoders. Besides, that is particularly created to recognize the impact of time on repetitions in sequenced data hence, it can be used to perform time series analysis and anomaly detection. Moreover, LSTM analysis can be deliberated using eqn. (14),

$$V_x = \delta(u_q \cdot (H_{(T-1)}, P_x) + B_q) \quad (14)$$

What δ is (activation function), $H_{(T-1)}$ denotes the last hidden state vector and P_x defines the input.

Encoder: Hidden Markov networks have an approach to deal with sequential input (for example, time series data or event sequences) using the encoder that is based on Long Short-Term Memory units. But these units are an efficient kind of recurrent neural network that can model the long-range dependencies in sequential data in such a way that it maintains a cell state and controls the information flow through gates.

Latent Space Representation: The LSTM Autoencoder through the dimension reduction process utilizes the lower-dimensional representation of the input sequence, which is called the latent space or encoding. This model compresses the original sequence into a few vector representations while preserving the essential features and temporal dependencies.

Decoder: The disguised representation then enters the decoder which is constructed of LSTM units. The task of the decoder is to recover the original sequence from the compressed representation, in which the decoder aims to get as close as possible to the similarity between the reconstructed output and original input. During training, the objective to minimize reconstruction loss, which is a function that measures the differences between input sequences and the reconstructions obtained from them, is set. After training, the anomalies can be detected based on reconstruction errors; large reconstruction errors imply that the learned temporal patterns of an instance are significantly abnormal, and hence such instances are treated as anomalies.

3.8.2. Meta-Classifiers:

(i) XGBoost

In the stacking ensemble model setting, XGBoost acts as a meta-classifier responsible for the coordination of the integration of predictions from different base classifiers to improve model performance. XGBoost uses the process of eXtreme Gradient Boosting to learn how to combine the outputs of base classifiers during training and consequently assign different weights to each classifier's prediction based on its contribution to overall predictive accuracy. To make the ensemble strong, XGBoost has incorporated the regularization techniques that guarantee against overfitting and help the algorithm to perform better on new test data. It is computationally efficient, hence it is easy to train, and deploy with minimal resource needs, especially, when working with several base classifiers. Moreover, XGBoost includes interpretability, such as insights into feature importance, along with the stacking ensemble framework and therefore enhances transparency in decision-making. Ultimately, having a variety of working mechanisms and high efficiency, XGBoost becomes a leading classification meta-learner in the stacking ensembles that contribute to the increase of predictive power and robustness. The classification by XG boost classifiers can be deliberated using eqn. (15)

$$Q^{(n)} = \sum_{k=1}^t Q(Z_k, Z_k^{(n-1)} + M_n(u_k)) + \mu(M_n) \quad (15)$$

Where, $\sum_{k=1}^t Q(Z_k, Z_k^{(n-1)} + M_n(u_k)) = F(u + \Delta u)$. Moreover, the value of parameter u is, $u = Z_k^{(n-1)}$ respectively.

Comment: Can the operations performed be shown in algorithmic representation from subsection 3.5

4. Experimental Results

This study used Keras and the TensorFlow deep learning package on a 64-bit Intel Core-i7 CPU running Windows 7 with 16 GB of RAM. The machine learning method was executed in MATLAB. The CICIDS2017 dataset is used to evaluate the IDS approach, with 80% of the data being utilized for training and 20% for testing.

DATASET DESCRIPTION

CICIDS2017 dataset, which mimics actual real-world data, including safe and current common assaults (PCAPs). Along with labelled flows based on time stamps, source and destination IP addresses, source and destination ports, protocols, and attacks, it also contains the findings of a network traffic analysis performed with a CIC Flow Meter (CSV files). The definition of extracted characteristics is also accessible. Moreover, the samples present in the selected dataset can be tabulated in Table 1.

Table 1: Samples present in the dataset

Day	Description	Size (GB)
Monday	Normal Activity	11.0
Tuesday	Attacks + Normal activity	11G
Wednesday	Attacks + Normal Activity	13G
Thursday	Attacks + Normal Activity	7.8G
Friday	Attacks + Normal Activity	8.3G

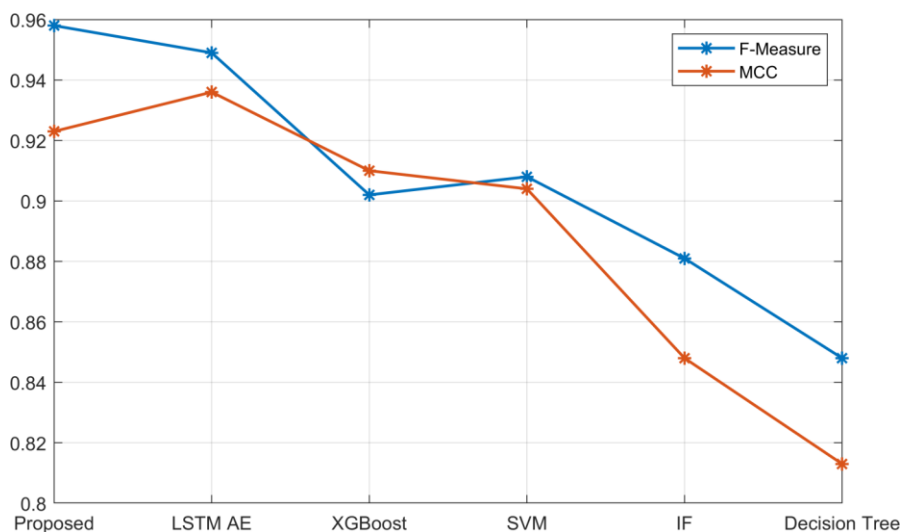


Figure 2: f-measure and MCC analysis

From Figure 2, it can be noted that the F-measure of the proposed MedSecEnsemble model is high with a value of 0.958. The F-measure of the existing methods like LSTM AE, XGBoost, SVM, IF, and Decision Tree are 0.949, 0.902, 0.908, 0.881, and 0.848 respectively. Similarly, the analysis from Fig. 2 shows that the MCC of the proposed MedSecEnsemble model is high (0.923) compared to the existing methods. The MCC of LSTM AE is 0.936, XGBoost is 0.910, SVM is 0.904, IF is 0.848, and Decision Tree is 0.813. These results show the efficiency of the proposed MedSecEnsemble model in improving the F-measure and MCC metric.

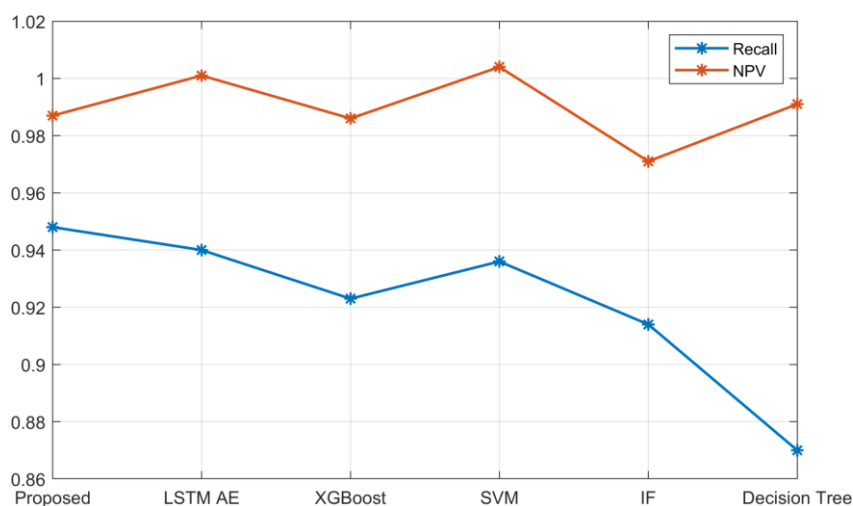


Figure 3: Recall and NPV analysis

With a recall rate of 0.948, the proposed MedSecEnsemble approach performs exceptionally well, showing that it can recover a high proportion of true positive events on the dataset. On the contrary, LSTM AE has recalls of 0.940, and XGBoost has recalls of 0.923. The SVM and IF methods are the next two with 0.936 and 0.914, respectively, in terms of their reasonable recall values. Of all models, the Decision Tree one with a recall of 0.870 is the poorest in terms of locating positive cases in the dataset. This NPV analysis clinches the ensemble method's accuracy in rightfully labelling true negatives, with LSTM AE and XGBoost ranking next. The SVM and IF show NPVs of 1.004 and 0.971 respectively and the Decision Tree model presents with an NPV of 0.991.

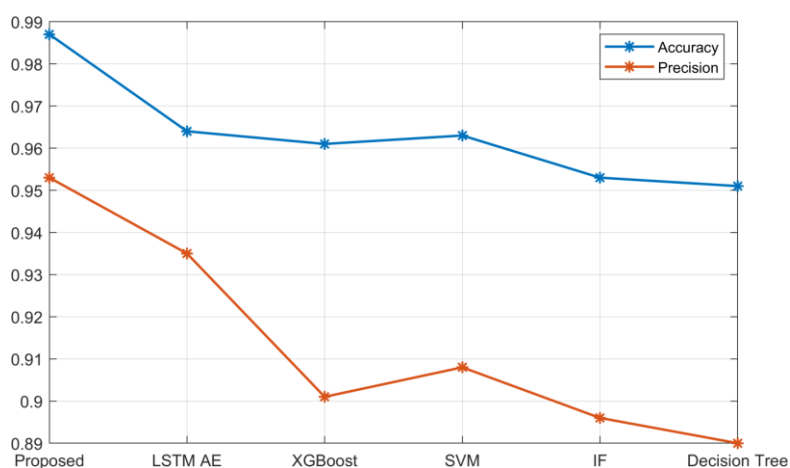


Figure 4: Accuracy and Precision Analysis

In Fig.4, the proposed MedSecEnsemble technique shows a high accuracy of 0.97, suggesting its good fetching inaccurate prediction. This exhibits that the combination method can make an overall balance between true positives, true negatives, false positives, and false negatives and therefore is a robust choice in the context of personal data protection. Furthermore, the XGBoost model performs very well with an accuracy of 0.961 whereas the Decision Tree model demonstrates a lower accuracy level of 0.951. The fact that the model has a higher precision shows its ability to minimize the number of false positives. The decision Tree model manifests with a precision of 0.890, and the SVM and IF models lead to higher levels of precision in the amount of 0.908 and 0.896, respectively.

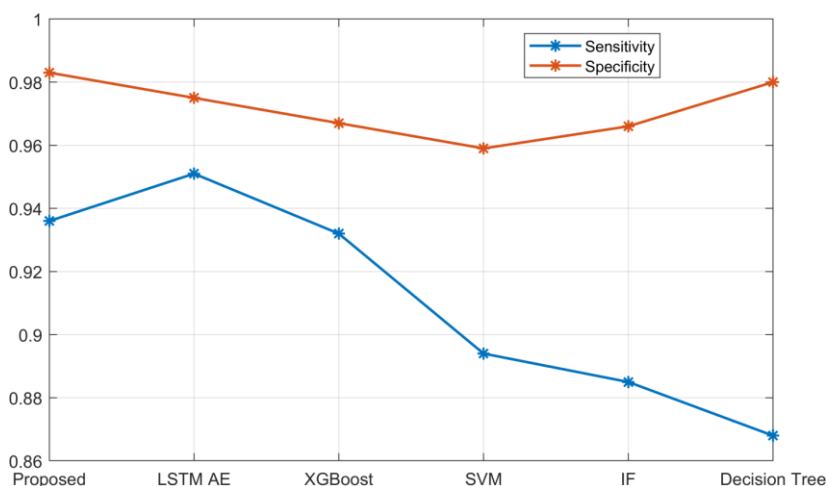


Figure 5: Sensitivity and Specificity Analysis

The depicted MedSecEnsemble technique in Figure 5 has a sensitivity of 0.948 which is higher proof of its ability to correctly identify the real positive instances. However, the Decision Tree has a lower sensitivity of 0.870. The proposed ensemble approach shows very high specificity (0.983), using LSTM AE returns slightly less specificity (0.975), and the XGBoost model also performs well with a specificity of 0.967. However, the SVM and IF display specificities of 0.959 and 0.966, respectively. The results of the study are very significant in data protection of personal type where reducing the false positives is a must for the integrity of the security system and the absence of useless interventions.

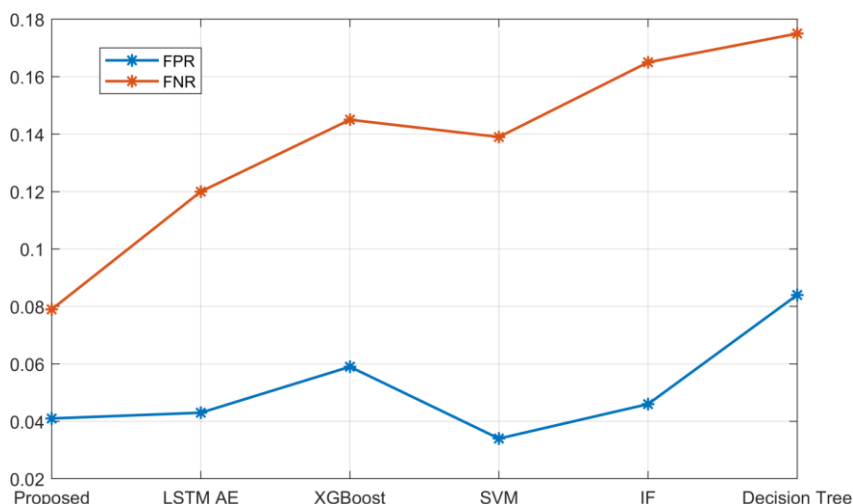


Figure 6: FPR and FNR analysis

The proposed MedSecEnsemble technology, with an FPR of 0.041, shows the potential of keeping the number of false alarms to a minimum. The LSTM AE has a slightly higher FPR of 0.043. The XGBoost model also performs well with an FPR of 0.059. The SVM and IF demonstrate FPRs of 0.034 and 0.046, respectively whereas, the Decision Tree model exhibits a higher FPR of 0.084, suggesting a slightly higher rate of misclassifying negative instances. Similarly, the FNR is also found to be low for the proposed MedSecEnsemble technique compared to other existing techniques. Table 2 provides the overall summary of the comparative analysis with numerical results.

Table 2: Performance analysis of the proposed MedSecEnsemble technique

Methods/ Metrics	Sen	Spec	Acc	Precisi on	Reca ll	FMeasu re	NP V	FP R	FN R	MCC
Proposed	0.936	0.983	0.987	0.953	0.948	0.958	0.987	0.041	0.079	0.923

LSTM AE	0.951	0.975	0.964	0.935	0.94	0.949	1.001	0.043	0.12	0.936
XGBoost	0.932	0.967	0.961	0.901	0.923	0.902	0.986	0.059	0.145	0.91
SVM	0.894	0.959	0.963	0.908	0.936	0.908	1.004	0.034	0.139	0.904
IF	0.885	0.966	0.953	0.896	0.914	0.881	0.971	0.046	0.165	0.848
Decision Tree	0.868	0.98	0.951	0.89	0.87	0.848	0.991	0.084	0.175	0.813

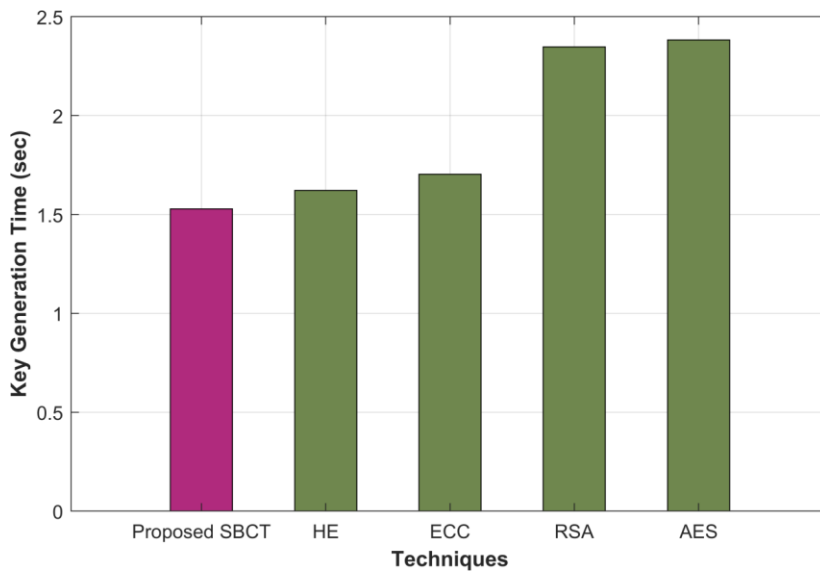


Figure 8: Key Generation time analysis

The proposed SBCT method has a key generation time of 1.528 sec which is comparatively lower than existing methods such as HE (1.622 sec), ECC (1.703 sec), RSA (2.374 sec), and AES (2.383 sec). Among the existing methods, AES has taken the highest amount of time for key generation. The reason for the improvement in the proposed method is attributed to the proposed ETDO algorithm.

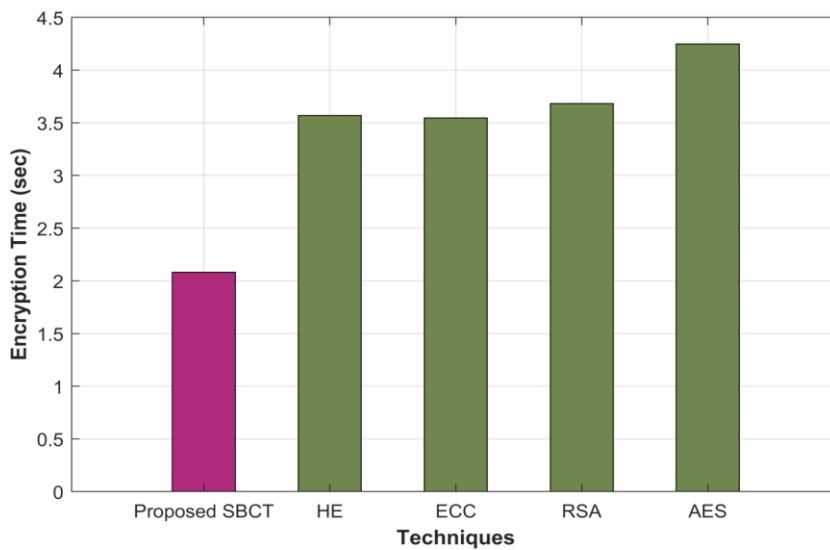


Figure 9: Encryption time analysis

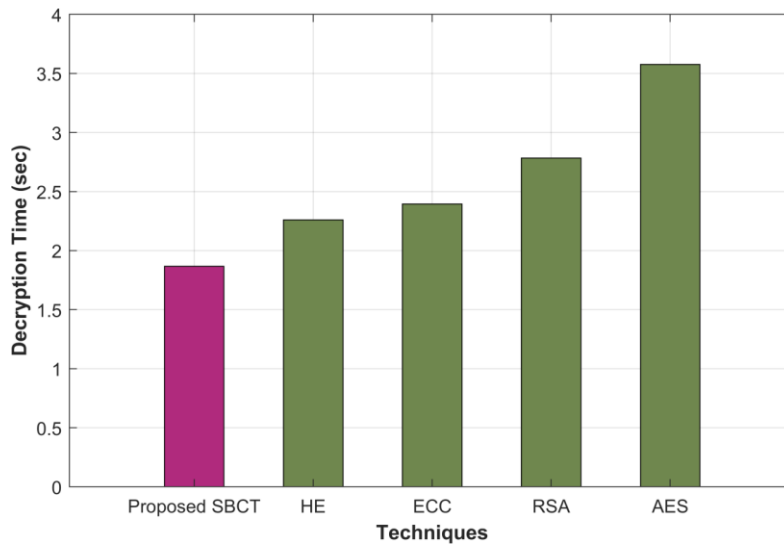


Figure 10: Decryption time analysis

The encryption times for several encryption techniques are shown in Figure 9. With an encryption time of 2.081 seconds, the suggested SBCT technique is the fastest, followed by HE at 3.57 seconds, ECC at 3.546 seconds, RSA at 3.682 seconds, and AES at 4.249 seconds. This suggests that the proposed SBCT method is the most efficient in terms of encryption time compared to the other methods.

The decryption time of several encryption algorithms as represented in Fig 10 reveals that the SBCT method has the edge over all the other methods specifically in terms of the speed of decryption. The decryption time of 1.868 seconds is much better than what is the case with other broadly used algorithms. This performance edge is vital to take into account mainly in those circumstances where rapid decryption is required as, for instance, in real-time data processing and communication systems.

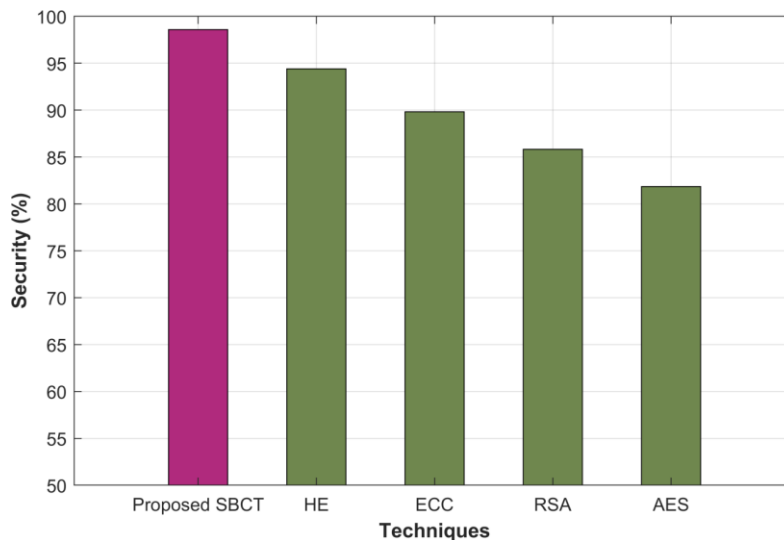


Figure 11: Security analysis

Figure 11 demonstrates the effectiveness of the Proposed SBCT algorithm in maintaining a high level of data security. Homomorphic Encryption (HE) ensures a robust security level of 94.4%. ECC provides a security level of 89.818% along with moderate encryption and decryption times. Among the compared methods, AES exhibits a lower security of 81.86%. Detailed numerical results can be found in Table 2.

Table 2: Performance analysis of the proposed SBCT Encryption algorithm

Method/Metrics	Encryption Time (sec)	Decryption Time (sec)	Key generation Time (sec)	Security (%)
Proposed SBCT	2.081	1.868	0.001	98.18
HE	3.57	2.27	0.001	94.4
ECC	3.546	2.4	0.001	89.818
RSA	3.682	2.8	0.001	86
AES	4.249	3.57	0.001	81.86

Proposed SBCT	2.081	1.868	1.528	98.595
HE	3.57	2.259	1.622	94.4
ECC	3.546	2.394	1.703	89.818
RSA	3.682	2.784	2.347	85.829
AES	4.249	3.577	2.383	81.86

Table 3: Accuracy comparison

Methods [24]	Accuracy
SVM (2012)	82.05%
CNN (2017)	46.97%
MLP (2018)	90%
Word2vev (2018)	83%
KNN (2018)	97%
Proposed	98.7%

5. Conclusions

In the proposed paper, a full-scale method named PriMedGuard is presented for the protection of personal medical data. In the beginning, the methods involve the data collection and then, robust pre-processing techniques like the missing values handling and normalization follow this step. This ensures that the data collected is reliable, coherent, and ready for exploration. The addition of the enhanced Tasmanian Devil Optimization Algorithm for key generation reinforces the privacy and consistency of medical information. The SBCT algorithm is used in the data encryption process that allows secure computation of the encrypted data, which doesn't require any decryption. This advanced encryption method is a critical element of medical data security especially while it is transmitted and stored in a decentralized IPFS. Further, the machine learning approach of intrusion detection for the IoMT network, that is, the MedSecEnsemble technique, is attributed to the continuous surveillance and speedy recognition of the potential security breaches thus making the method more robust.

Comment: Add the following comparison in the form of Table

1. Privacy rights and data security: GDPR and personal data markets. *Management Science*, 69(8), 4389-4412.
2. Secure health data sharing for medical cyber-physical systems for the healthcare 4.0. *IEEE journal of biomedical and health informatics*, 24(9), 2499-2505.
3. Hernández-Álvarez, L., de Fuentes, J. M., González-Manzano, L., & Hernández Encinas, L. (2020). Privacy-preserving sensor-based continuous authentication and user profiling: a review. *Sensors*, 21(1), 92. (USE TABLE TO SHOW Comparison)
4. Cost-effective authenticated data redaction with privacy protection in IoT. *IEEE Internet of Things Journal*, 8(14), 11678-11689.
5. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications*, 50, 102407.
6. COVID-19, digital privacy, and the social limits on data-focused public health responses. *International Journal of Information Management*, 55, 102181.
7. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6), 305-311.
8. Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics*, 22, 1-5.

References

- [1] Kumar P, Kumar R, Srivastava G, Gupta GP, Tripathi R, Gadekallu TR, Xiong NN. PPSF: A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. *IEEE Transactions on Network Science and Engineering*. 2021 Jun 16;8(3):2326-41.
- [2] Elhoseny M, Haseeb K, Shah AA, Ahmad I, Jan Z, Alghamdi MI. IoT solution for AI-enabled PRIVACY-PREServing with big data transferring: an application for healthcare using blockchain. *Energies*. 2021 Aug 28;14(17):5364.
- [3] Kumar R, Wang W, Kumar J, Yang T, Khan A, Ali W, Ali I. An integration of blockchain and AI for secure data sharing and detection of CT images for the hospitals. *Computerized Medical Imaging and Graphics*. 2021 Jan 1; 87:101812.
- [4] Haddad A, Habaebi MH, Islam MR, Hasbullah NF, Zabidi SA. Systematic review on ai-blockchain based e-healthcare records management systems. *IEEE Access*. 2022 Aug 26; 10:94583-615.
- [5] Sivan R, Zukarnain ZA. Security and privacy in cloud-based e-health system. *Symmetry*. 2021 Apr 23;13(5):742.
- [6] Pise AA, Almuzaini KK, Ahanger TA, Farouk A, Pareek PK, Nuagah SJ. Enabling artificial intelligence of things (AIoT) healthcare architectures and listing security issues. *Computational Intelligence and Neuroscience*. 2022 Aug 3;2022.
- [7] Thapa C, Camtepe S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Computers in biology and medicine*. 2021 Feb 1; 129:104130.
- [8] Garg N, Petwal R, Wazid M, Singh DP, Das AK, Rodrigues JJ. On the design of an AI-driven secure communication scheme for Internet of medical things environment. *Digital Communications and Networks*. 2023 Oct 1;9(5):1080-9.
- [9] Ramasamy LK, Khan F, Shah M, Prasad BV, Iwendi C, Biamba C. Secure smart wearable computing through artificial intelligence-enabled internet of things and cyber-physical systems for health monitoring. *Sensors*. 2022 Jan 29;22(3):1076.
- [10] Liu Y, Yu J, Fan J, Vijayakumar P, Chang V. Achieving privacy-preserving DSSE for intelligent IoT healthcare system. *IEEE Transactions on Industrial Informatics*. 2021 Jul 30;18(3):2010-20.
- [11] Lo'ai AT, Saldamli G. Reconsidering big data security and privacy in cloud and mobile cloud systems. *Journal of King Saud University-Computer and Information Sciences*. 2021 Sep 1;33(7):810-9.
- [12] Ali A, Rahim HA, Pasha MF, Dowsley R, Masud M, Ali J, Baz M. Security, privacy, and reliability in digital healthcare systems using blockchain. *Electronics*. 2021 Aug 23;10(16):2034.
- [13] Rahman A, Hossain MS, Muhammad G, Kundu D, Debnath T, Rahman M, Khan MS, Tiwari P, Band SS. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Cluster computing*. 2023 Aug;26(4):2271-311.
- [14] Rehman A, Saba T, Haseeb K, Larabi Marie-Sainte S, Lloret J. Energy-efficient IoT e-health using artificial intelligence model with homomorphic secret sharing. *Energies*. 2021 Oct 7;14(19):6414.
- [15] Andreas A, Mavromoustakis CX, Mastorakis G, Do DT, Batalla JM, Pallis E, Markakis EK. Towards an optimized security approach to IoT devices with confidential healthcare data exchange. *Multimedia Tools and Applications*. 2021 Aug; 80:31435-49.
- [16] Mahajan HB, Rashid AS, Junnarkar AA, Uke N, Deshpande SD, Futane PR, Alkhayyat A, Alhayani B. Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems. *Applied Nanoscience*. 2023 Mar;13(3):2329-42.
- [17] Ali A, Pasha MF, Ali J, Fang OH, Masud M, Jurcut AD, Alzain MA. Deep learning based homomorphic secure search-able encryption for keyword search in blockchain healthcare system: A novel approach to cryptography. *Sensors*. 2022 Jan 11;22(2):528.
- [18] Sundas A, Badotra S, Bharany S, Almogren A, Tag-ElDin EM, Rehman AU. HealthGuard: An Intelligent Healthcare System Security Framework Based on Machine Learning. *Sustainability*. 2022 Sep 22;14(19):11934.
- [19] Mohanty MD, Das A, Mohanty MN, Altameem A, Nayak SR, Saudagar AK, Poonia RC. Design of Smart and Secured Healthcare Service Using Deep Learning with Modified SHA-256 Algorithm. *InHealthcare* 2022 Jul 9 (Vol. 10, No. 7, p. 1275). MDPI.
- [20] Kumar P, Kumar R, Gupta GP, Tripathi R, Jolfaei A, Islam AN. A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system. *Journal of Parallel and Distributed Computing*. 2023 Feb 1; 172:69-83.
- [21] Almalawi A, Khan AI, Alsolami F, Abushark YB, Alfakeeh AS. Managing Security of Healthcare Data for a Modern Healthcare System. *Sensors*. 2023 Mar 30;23(7):3612.

- [22] Ghazal TM, Hasan MK, Abdullah SN, Bakar KA, Al Hamadi H. Private blockchain-based encryption framework using computational intelligence approach. *Egyptian Informatics Journal*. 2022 Dec 1;23(4):69-75.
- [23] Ali A, Almaiah MA, Hajje F, Pasha MF, Fang OH, Khan R, Teo J, Zakarya M. An industrial IoT-based blockchain-enabled secure searchable encryption approach for healthcare systems using neural network. *Sensors*. 2022 Jan 12;22(2):572.
- [24] Hernández-Álvarez, L., de Fuentes, J. M., González-Manzano, L., & Hernández Encinas, L. (2020). Privacy-preserving sensor-based continuous authentication and user profiling: a review. *Sensors*, 21(1), 92.