Information Dynamics and Applications
https://www.acadlore.com/journals/IDA

Enhancing Healthcare Data Security in IoT Environments Using Blockchain and DCGRU with Twofish Encryption

Kumar Raja Depa Ramachandraiah[1], Naga Jagadesh Bommagani[2*], Praveen Kumar Jayapal[3]

[1] Faculty of Information and Communications Technology, Universiti Teknikal Malaysia Melaka, 76100 Melaka, Malaysia
[2] School of Computer Science and Engineering, VIT-AP University, 522237 Vijayawada, India
[3] DiSTAP, Singapore-MIT Alliance for Research and Technology, 138602 Singapore, Singapore

* Correspondence: Naga Jagadesh Bommagani (nagajagadesh@gmail.com)

Received: 10-11-2023 **Revised:** 11-15-2023 **Accepted:** 11-22-2023

Citation: K. R. D. Ramachandraiah, N. J. Bommagani, and P. K. Jayapal, "Enhancing healthcare data security in IoT environments using blockchain and DCGRU with twofish encryption," *Inf. Dyn. Appl.*, vol. 2, no. 4, pp. 173–185, 2023. https://doi.org/10.56578/ida020402.

Abstract: In the rapidly evolving landscape of digital healthcare, the integration of cloud computing, Internet of Things (IoT), and advanced computational methodologies such as machine learning and artificial intelligence (AI) has significantly enhanced early disease detection, accessibility, and diagnostic scope. However, this progression has concurrently elevated concerns regarding the safeguarding of sensitive patient data. Addressing this challenge, a novel secure healthcare system employing a blockchain-based IoT framework, augmented by deep learning and biomimetic algorithms, is presented. The initial phase encompasses a blockchain-facilitated mechanism for secure data storage, authentication of users, and prognostication of health status. Subsequently, the modified Jellyfish Search Optimization (JSO) algorithm is employed for optimal feature selection from datasets. A unique health status prediction model is introduced, leveraging a Deep Convolutional Gated Recurrent Unit (DCGRU) approach. This model ingeniously combines Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) processes, where the GRU network extracts pivotal directional characteristics, and the CNN architecture discerns complex interrelationships within the data. Security of the data management system is fortified through the implementation of the twofish encryption algorithm. The efficacy of the proposed model is rigorously evaluated using standard medical datasets, including Diabetes and EEG Eyestate, employing diverse performance metrics. Experimental results demonstrate the model's superiority over existing best practices, achieving a notable accuracy of 0.884. Furthermore, comparative analyses with the Advanced Encryption Standard (AES) and Elliptic Curve Cryptography (ECC) models reveal enhanced performance metrics, with the proposed model achieving a processing time and throughput of 40 and 45.42, respectively.

Keywords: Blockchain; Modified Jellyfish Search Optimization; Gated Recurrent Unit (GRU); Internet of Things (IoT); Data security; Twofish algorithm

1 Introduction

The advent of IoT has revolutionized the modern landscape by seamlessly integrating a myriad of devices with the internet, thereby enhancing life quality with minimal human intervention [1]. A significant acceleration in the proliferation of internet-enabled devices has been observed, with projections by the CISCO-2020 report indicating a requirement for over 50 billion connected devices. IoT stands as a testament to technological advancement, simplifying interactions with physical objects and finding applications in diverse sectors such as smart healthcare, smart agriculture, and smart urban planning [2]. The data generated by these IoT devices, encompassing both structured and unstructured forms, is expanding exponentially, necessitating robust measures for secure data management. The four Vs of big data, namely, velocity, volume, value, and veracity, underscore the imperative for stringent security protocols. In response to this, considerable research has been conducted to understand and categorize security assessment methodologies, aiming to bolster big data security services [3].

Blockchain technology emerges as a pivotal solution in this context, with its inherent immutability ensuring that once data is entered, it becomes impervious to unauthorized alterations. This feature is crucial in maintaining

the integrity of patient records and mitigating the risk of data manipulation. Additionally, blockchain employs cryptographic techniques to secure data, with each network participant possessing a unique private key, thereby encrypting transactions and safeguarding sensitive information from unauthorized access. Furthermore, blockchain facilitates meticulous access control via smart contracts, enabling patients to dictate access parameters to their health data, thereby enhancing both privacy and security.

The utility of healthcare data extends beyond mere record-keeping; it plays an instrumental role in disease prevention and treatment [4]. The rapid advancement of AI has rendered medical records an invaluable resource for developing AI-based diagnostic models, assisting healthcare professionals in clinical decision-making. Despite this progression, concerns regarding data privacy persist [5]. Several healthcare facilities have limited data transfer and sharing, resulting in patient records becoming fragmented across disparate data silos [6]. This fragmentation poses a significant challenge in the healthcare sector, as it hinders the holistic care of patients due to restricted data accessibility. A case in point is a notable breach in a major healthcare network, where unauthorized access to medical records led to the exposure of sensitive patient information. This incident not only breached patient confidentiality but also highlighted the pressing need for effective data protection mechanisms in the healthcare domain.

The challenges with data privacy and security in the healthcare sector extend beyond mere protection of information. For instance, patients visiting new hospitals often undergo redundant testing due to the lack of accessible medical history, leading to unnecessary financial and environmental burdens [7]. Furthermore, the inability to share sensitive patient data with research institutions hampers medical advancements. This predicament has catalyzed the exploration of secure data storage and transmission solutions, with blockchain's decentralized and tamper-proof nature emerging as a viable option for the safe transfer of medical information [8]. Innovations like Innoplexus harness AI and blockchain technology for real-time analysis of life science data globally, facilitating the flow of information to research institutions and pharmaceutical companies. Additionally, platforms such as BlockRx, integrating iSolve's advanced digital ledger technology with blockchain, amalgamate medical data from various biomedical and research institutes, demonstrating practical implementations of these technologies [9].

The integrity of patients' medical records is primarily threatened by issues in user-side verification, proper data querying [10], and the decentralized nature of blockchain technology, which offers enhanced security through digital signatures and encryption-based hashing algorithms. Given the high monetary value of personal information and medical records, the development of a reliable and secure data management infrastructure is imperative [11]. The reluctance of Cloud Service Providers (CSPs) to share medical records stems from potential risks associated with data leakage by attacker-controlled data users. To mitigate these risks, numerous cryptographic algorithms have been developed for secure healthcare data interchange and storage [12]. However, these methods often fall short in addressing user privacy needs and the failure potential of cloud servers. Encrypting data before transferring to internet-based servers is crucial, considering the variety of data generated in healthcare, including records, financial documents, clinical and imaging test results, and vital sign assessments [13].

Despite the complexities in accessing healthcare data and the existence of alternative data collection methods, the volume of databases in healthcare settings continues to grow. Blockchain technology holds promise in enhancing data verification and authenticity [14], facilitating information sharing across systems or services. These applications significantly impact the value, cost, and relevance of healthcare delivery. Being a distributed ledger, blockchain eliminates the need for multiple authentication layers in healthcare networks, making information readily accessible to all authorized users [15].

This study proposes an IoT-based, blockchain-driven, physics-inspired approach for healthcare delivery. The modified JSO algorithm is employed for feature selection, while the DCGRU model classifies health states into normal and pathological categories. The structure of the study is organized as follows: review of relevant literature in Section 2, detailed description of the proposed model in Section 3, presentation of results and discussion in Section 4, and summary and conclusions in Section 5.

2 Related Works

In the domain of Industrial Internet of Things (IIoT) network security, the work by Selvarajan et al. [16] has been instrumental. This study presents a novel lightweight intelligence framework, primarily designed to safeguard the confidentiality and integrity of IIoT networks. The core innovation of this research lies in its integration of lightweight blockchain with the Convivial Optimised Sprinter Neural Network (COSNN)-based AI processes. Utilizing the Authentic Intrinsic Analysis (AIA) model, the framework effectively translates characteristics into encoded data, thereby mitigating the impact of potential attacks. Rigorous testing on diverse attack datasets was conducted to ascertain the framework's effectiveness. Additionally, the performance of privacy safeguards and AI techniques was independently evaluated and compared using multiple metrics. Remarkably, the proposed AILBSM framework achieved a detection performance of 99.7%, reduced execution time to 0.6 seconds, and enhanced classification accuracy to 99.8%. These results indicate a substantial improvement in anomaly detection capabilities over existing methods.

Another significant contribution to blockchain-based security parameter categorization is provided by Kamalov et al. [17]. This research delves into identifying the fundamental causes of security challenges using a laboratory method. The use of fuzzy logic, renowned for addressing the inherent ambiguity in human judgment, is a focal point of this study. The findings underscore the efficacy of the proposed methodology, contributing vital insights to the fields of confidentiality, integrity, and availability (CR1-3).

In the realm of smart healthcare, Alruwaill et al. [18] proposed a blockchain-based scheme employing Internet of Medical Things (IoMT) devices for real-time patient monitoring. This scheme includes data hashing and encryption at the edge device level, coupled with additional processing capabilities. The use of blockchain technology in this context ensures robust protection against unauthorized access, facilitated by symmetric key encryption. Patients are thus enabled to securely transmit data to healthcare providers via smart contracts. In the healthcare provider's system, the integration of a verification node and blockchain, utilizing an asymmetric key, provides a mechanism for signing and confirming the authenticity and origin of patient data. The study also examines location-based authentication as a means to validate data veracity and provenance.

Li et al. [19] have developed a blockchain-based method, termed BFOD, for the exchange of flight operating data, aimed at enhancing privacy and secure data sharing. In this model, the participants are categorized into data owners, data requesters, and authorization institutes, in line with the business logic. Each civil aviation company is assigned a hash anonymous identity and a specific set of permissions by the authorization institution, ensuring confidentiality. The validation of data requesters' credentials is performed using a hash anonymous identity, safeguarding the identities of data users. Furthermore, the authenticity of the flight operation data is verified using zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs), which confirms compliance with the data requesters' requirements without compromising confidentiality. A proxy re-encryption method is employed to enhance the efficiency of flight operations data exchange. Additionally, a grouped Practical Byzantine Fault-Tolerant (PBFT) approach is proposed to reduce consensus delay. Theoretical and experimental results indicate that the grouped PBFT algorithm, with a set number of consensus nodes (N=100), reduces consensus latency by 91.4%. The BFOD is thereby demonstrated to be practical, efficient, and protective of user data privacy.

Fang et al. [20] addressed the challenges in the blockchain-based global organization of IoT. The study begins with an analysis of the structural models of IoT and blockchain technology. Subsequently, it investigates issues related to zero-knowledge proof (ZKP) and the trusted execution environment (TEE) in blockchain-based information security assurance. A system is developed to safeguard the security and privacy of stored data using blockchain technology. The feasibility of the proposed system is validated through simulated experiments. According to experimental findings from the top node (N=28), the recommended methodology generates proof within a mere 352 milliseconds, marking a substantial improvement over previous methods.

In the context of IoT-based healthcare applications, Ali et al. [21] proposed an innovative approach to enhance privacy protection using homomorphic encryption methods and blockchain technology. Homomorphic encryption enables computations on encrypted data without decryption, thus maintaining data confidentiality during processing. The proposed method employs smart contracts within the blockchain network to enforce access restrictions and establish data-sharing rules. These smart contracts feature granular permission controls, limiting the use of encrypted information to authorized parties only. The approach also generates an audit trail of all data transactions, enhancing transparency and accountability. Comparative analysis with conventional models has been conducted, considering communication costs, transaction volumes, and security levels. The trials demonstrate the effectiveness of this method in preserving data privacy while allowing for efficient data processing and analysis. In summary, the combination of homomorphic encryption and blockchain technology offers a robust solution for protecting patient privacy in healthcare IoT applications.

Dewangan et al. [22] propose a novel system employing the InterPlanetary File System (IPFS), along with the Edward- and Secure Hash Algorithm (SHA-256) for digital signature and verification. The system uses tokens to construct student identities, which are then stored in the IPFS. The performance of the proposed system is evaluated based on the time taken for signing and verifying a transaction, as well as the total transaction completion time. A comparative analysis is conducted, considering aspects such as privacy, transaction costs, large file storage, blockchain implementation, and registration costs, demonstrating the efficacy of the proposed system.

3 Proposed Methodology

In the envisaged healthcare IoT framework, devices such as thermostats and intrusion detection systems are integrated with servers. The methodology proposes the interconnection of multiple states using blockchain technology to create a comprehensive medical data hub, incorporating end IoT devices. This study considers practical scenarios where users, with appropriate permissions, access and manage data. For instance, an individual named Alice can remotely view temperature data from her IoT devices. The methodology involves expanding the cryptographic key using the twofish key expansion algorithm, which transforms the original key into an array of subkeys for encryption and decryption. Subsequently, the twofish encryption algorithm is applied to each data block

using the expanded key, involving Feistel network rounds, key mixing, S-box operations, and linear transformations.

3.1 Registration Phase

In the registration phase, hospital staff inputs patient data onto the cloud server. Upon successful registration, patients are assigned a unique ID. A hash code, generated using the Adler-32 hashing method, combines the patient's unique identifier, registration time, and other relevant data. This hash code, utilized for user identity verification during login, is calculated through a 32-bit hash comprising two 16-bit checksums, *G* and *H*. The Adler-32 hashing technique computes H as the cumulative sum of individual G values, both starting at Eq. (1). The patient's identifier and registration time are denoted by β and *S* in the methodology, respectively. The hash codes are formulated by amalgamating the patient's identifier with the current time.

$$\alpha = \beta S \tag{1}$$

The Adler-32 hashing algorithm then generates the hash code for this combined data. In this algorithm, a prime number, specifically 65521.2^{16} (the largest prime number less than 65521), is employed.

$$H = 1 + \alpha_1 + \alpha_2 \ldots\ldots\ldots + \alpha_m \cdot |65521| \tag{2}$$

In the following equation, *G* represents the sum of each value of *H*, and *m* is the length of the data string α.

$$G = \langle (1 + \alpha_1) + (1 + \alpha_1 + \alpha_2) + \cdots + (1 + \alpha_2 + \cdots + \alpha_m) \rangle \cdot |65521| \tag{3}$$

The hash code generation process is detailed in the subsequent equation:

$$\mathrm{Adl}(\alpha) = (G \times 65521) + H \tag{4}$$

The hash code $Adl(\alpha)$ is provided to the patient upon registration. The patient's data, once registered, is securely and confidentially stored on a distributed ledger, known as a blockchain.

3.2 Blockchain

With the expansion of the healthcare sector, ensuring patient confidentiality while managing health records has emerged as a formidable challenge. Blockchain has been proposed as a solution to these issues, facilitating the secure transmission of medical records. Blockchain, essentially a distributed database, records all transactions in a public, immutable digital ledger comprising a cryptographically linked series of blocks. Once added to the blockchain ledger, data blocks become immutable, precluding any deletion. The majority of data within blockchain is stored in an encrypted linked list of transactions. This approach employs the SHA-256 for encrypting input data and generating hash results, thus revealing any alterations made to a block. The blockchain employs a proof-of-work (POW) system to establish reliable consensus mechanisms. Altering the blockchain is computationally intensive, thereby deterring malicious attacks. Adding a new block to the network requires solving a POW. The block header records the block size, the Merkle root, the blockchain version, and the hash code of the preceding block. The Merkle root is the central hash value in the Merkle tree, with the remaining transactions structured in this tree-like formation. The block header also includes the current date, the difficulty target (regarding POW), and the nonce, a random value used in determining the POW.

3.3 User Authentication

During user authentication, the data provided at registration and login are compared. Authentication is confirmed when both the fingerprint and the data correspond. A mismatch triggers a warning, and data remains inaccessible until successful authentication. Upon verification of a patient, health data can be collected using wireless body sensors (WBS). While WBS enhances the quality of life, it does not inherently protect user data privacy during wireless transmission. Consequently, it is imperative to encrypt data during transmission to ensure privacy and security.

3.4 Data Encryption Utilizing the Twofish Algorithm

In the realm of IoT healthcare, safeguarding the integrity of data collected at the sensor layer is paramount. For this purpose, the twofish algorithm has been employed to encrypt the gathered data. The appeal of the twofish algorithm lies in its remarkable efficiency and versatility, making it suitable for various platforms. It features an encryption block size of 128 bits and supports key lengths up to 256 bits, thereby expanding its applicability and enhancing security [23]. The versatility of twofish allows it to operate effectively across a range of devices, from a 32-bit central processing unit (CPU) to an 8-bit smart card, and can be adapted for both small and large-scale circuits. Its architecture supports diverse performance levels, influenced by key size, memory requirements,

hardware specifications, and encryption operation speed. The efficacy of twofish in cryptographic applications is well-documented [23].

The encryption process involves receiving plaintext and applying the twofish method over 16 rounds, using 32-bit words as input. Prior to submitting feedback through the interface layer, client verification is crucial. As customer feedback is integral to the decision-making layer, it undergoes stringent verification before granting access. The proposed model facilitates user/customer validation, crucial for authentication. User authentication combines standard alphanumeric passwords, graphical pattern image creation, and biometric methods such as iris, fingerprint, and retina scans. The choice of biometric authentication method depends on system requirements. The comprehensive user/customer authentication process is delineated in Algorithm 1, which outlines the hybrid authentication procedure.

Algorithm 1: Hybrid user authentication procedure

1: Procedure INITIAL LOGIN PROCESS (*Userid*)
2: $U_{id} \rightarrow$ input username ▷ User enters a username
3: $U_{pass} \rightarrow$ input password ▷ User enters a pre-selected alphanumerical password
4: If $(U_{id} \&\& U_{pass}) ==$ True then
5: Go to the next practice;
6: Else
7: Discard process: ▷ Discard the process
8: AUTHENTICATION $\left(D_{trust}^{n-id}\right)$
9: $U_{fr1} \rightarrow$ Select first-round image ▷ Image selection from a set
10: $U_{fr2} \rightarrow$ Select second-round image
11: $U_{fr3} \rightarrow$ Select second-round image
12: If $(U_{fr1} \&\& U_{fr2} \&\& U_{fr3}) ==$ True then ▷ Decision making
13: Go to the next procedure
14: Else
15: Discard procedure: ▷ Discard the process
16: Procedure Biometric Authentication (U_{Bio}) ▷ Biometric Authentication procedure
17: $U_{ir} \rightarrow$ Iris scanning ▷ Iris based authentication
18: $U_{re} \rightarrow$ retina scanning ▷ Retina based verification
19: $U_{fp} \rightarrow$ Fingerprint scanning ▷ Finger based authentication
20: If $(U_{ir} \| U_{re} \| U_{fp}) ==$ True then ▷ Final decision-making
21: Allow user to make changes;
22: Else
23: Discard authentication process: ▷ Unsuccessful verification
24: Exit

3.5 Disease Prediction Using DCGRU

In the context of disease prediction V_{s} using a network, a function f is defined to represent the relationship between well-log data X and the target variable Y_{d} at depth d, as shown in Eq. (5):

$$Y_d = f(X, d) \tag{5}$$

Given the highly nonlinear nature of f, it is challenging to derive explicitly. Therefore, it is more feasible to approximate f using a Deep Neural Network (DNN). DNNs are capable of estimating Y_{d} by analyzing the trends in data and mapping them accordingly, as expressed in Eq. (6):

$$f_\theta : X, d \rightarrow Y \tag{6}$$

Eqs. (5) and (6) indicate that the DNN can be trained to approximate the mapping to V_{s}. The training process involves optimizing the network parameters to achieve this approximation.

For a series of data, the depth arrangement matrix X is formulated in Eq. (7). The set $X_d = (x_1^m, x_2^m, \ldots, x_d^m, \ldots, x_D^m)$ comprises values of depth d, and the set $X^m = (x_1^m, x_2^m, \ldots, x_d^m, \ldots, x_D^m)$ contains values at these depths, such as

V_{p}. The term $Y = (y_{a1}, y_{a2}, \ldots, y_{ad}, \ldots, y_{aD})$ represents the historical information of the predicted V_s :

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^m \\ x_2^1 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \vdots \\ x_d^1 & x_d^2 & \cdots & x_d^m \\ \vdots & \vdots & \vdots & \vdots \\ x_D^1 & x_D^2 & \cdots & x_D^m \end{bmatrix} \tag{7}$$

3.5.1 Network structure

The basic structure of the GRU model, as illustrated in Figure 1, is characterized by its focus on local connections. CNNs, particularly effective in computer vision applications, emphasize the correlation between neighboring pixels, which diminishes with increasing distance.

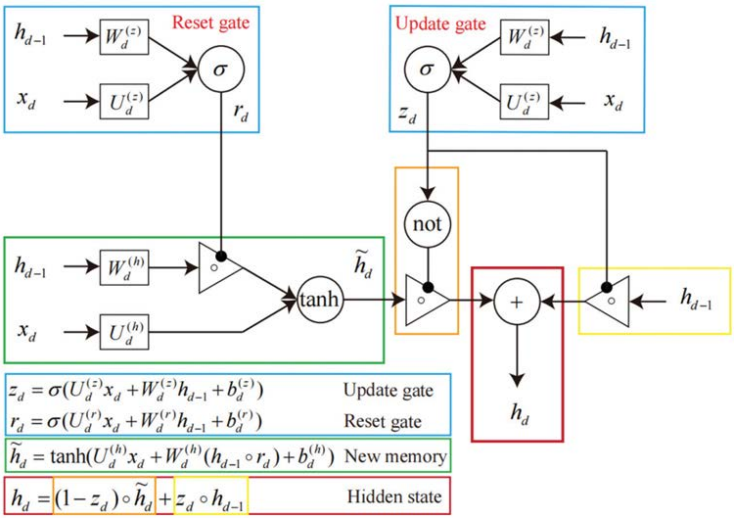

Figure 1. Internal composition of a GRU cell

The GRU network employs a gating mechanism to regulate input and deep memory data for prediction at a specific depth. CNNs are especially suitable for image processing, aligning with the principle of local connectivity. In well-log data, similar depths tend to exhibit similar information, making CNNs apt for feature extraction needed for prediction. The GRU network retains information of shallow layers, such as V_s values, utilizing it to predict deeper layers.

The DCGRU network, depicted in Figure 2, comprises an input layer, a CNN module, and a connected layer. The input module receives standard logging information, while the CNN module captures deep spatial properties of the logging data sequence. The GRU component processes depth-oriented well-log data, extracting essential features and capturing internal correlations. The final V_s prediction is then made in the fully connected layer.

3.5.2 Prediction framework

A sensitivity analysis is first conducted to compare conventional well logs with V_s for effective disease prediction. This involves constructing a data set from well-log data to evaluate the accuracy of the DCGRU model. The predictions from the DCGRU model are fed into the fully connected layer for V_s forecasting. Multiple iterations of training the DCGRU network may yield an optimal model. The best-performing DCGRU network is then employed to predict V_s, using root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) as performance metrics.

3.6 Feature Selection

Feature selection is an essential component in the deep learning framework, involving the identification of attributes that significantly contribute to the prediction of the unknown parameter. The effectiveness of prediction is influenced by both the quantity and quality of features selected. In this study, the DCGRU model's predictions are closely aligned with well-log data, underscoring the importance of well logs in prediction outcomes. Patient data, securely stored in the cloud, is accessible via well logs. Therefore, a preliminary analysis of the correlations between well logs and V_s is imperative prior to actual prediction. The modified JSO algorithm is utilized for feature selection.

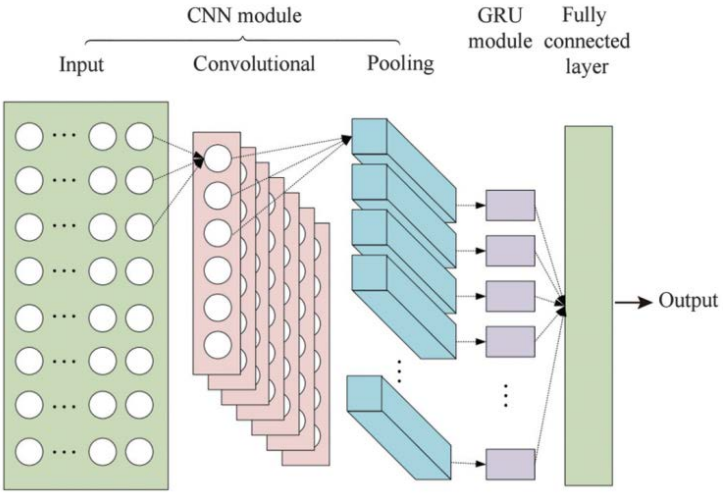

Figure 2. Architecture of the DCGRU fusion network

3.6.1 JSO

Jellyfish movement can be categorized into three distinct types. The first involves movement in the direction of the ocean current. The second type consists of jellyfish swimming within their swarm. The transition between these behaviors is governed by a time control (TC) procedure. A third type is characterized by jellyfish moving towards areas abundant in food, evaluated by a numerical fitness function.

The initial population of jellyfish can be mathematically represented as:

$$JF_i(It+1) = 4Y_0\left(1 - JF_i\right), 0 \le Y_0 \le 1, j = 1, 2, \ldots, N_{JF} \tag{8}$$

where, $JF_i(It+1)$ denotes jellyfish i, it is the iteration count, and Y_o indicates a random value within the range $[0, 1]$, excluding specific values of $\{0.0, 0.25, 0.75, 0.5, 1.0\}$. N_{JF} represents the total population of the jellyfish school. The TC procedure's value at each iteration (It) gradually transitions from 0 to 1 , as described by:

$$TC(It) = \left[\left(1 - \frac{It}{It^{\max}}\right) \times (-1 + 2 \times z_2)\right] \tag{9}$$

where, $It^{\max}$ is the maximum iteration count, and z_2 is a random number between 0 and 1.

If $TC(It)$ is greater than 0.5, the current jellyfish harvest site is determined by:

$$JF_i(It+1) = Z \times (JF_{\text{Best}} - 3 \times Z \times AV) + JF_i(It) \tag{10}$$

where, Z is a random number between 0 and 1, JF_{Best} represents the optimal location, and AV is the average location.

If $TC(It)$ is less than 0.5, jellyfish positions within the swarm are updated passively by:

$$JF_i(It+1) = 0.1 \times Z \times (U_b - L_b) + JF_i(It) \tag{11}$$

where, U_b and L_b are the limits of the regulating variables, respectively.

Alternatively, jellyfish may swim actively with the swarm as:

$$JF_i(It+1) = \begin{cases} JF_i(It) + \mathrm{R}\left(JF_j(It) - JF_i(It)\right) & \text{iff } (JF_i) \ge f\left(JF_j\right) \\ JF_i(It) + \mathrm{R}\left(JF_j(It) - JF_j(It)\right) & \text{iff } (JF_i) < f\left(JF_j\right) \end{cases} \tag{12}$$

The optimization problem's fitness function, representing the available food supply, is denoted by f.

During each iteration It, the following update is performed on each control variable in the jellyfish position vector JF_i :

$$JF(It)_{i,d} = U_{b,d} \text{ or } L_{b,d} \text{ if } JF(It)_{id} > U_{b,d} \text{ or } JF(It)_{i,d} < L_{b,d} \text{ respectively} \tag{13}$$

where, d is the index of a control variable within the jellyfish position vector.

3.6.2 Proposed modified JSO

In JSO, the balance between exploration and exploitation is meticulously maintained. To augment the JFS's search capabilities, the modified JSO incorporates a quasi-oppositional-based learning approach, thereby enhancing the exploration potential. This learning mechanism is integrated into the JSO and is applied to a subset of randomly selected jellyfish before testing the boundary limits as per Eq. (13):

$$JF_i(It+1) = \text{rand}\left[\frac{1}{2}\left(L_b + U_b\right), \left(L_b + U_b\right) - JF_i(It+1)\right] \text{ if } RA_2 < RA_3 \tag{14}$$

This procedure generates a quasi-opposite position for the jellyfish. Consequently, fitness calculations within the JSO remain consistent. Additionally, a random-sized social neighborhood group is constructed, facilitating effective intelligence exchange within this social context. The size of each jellyfish's social neighborhood group is determined as follows:

$$SNG_i = [JF_A : JF_B]_{1 \times AZi} \tag{15}$$

where, SNG_i represents the social neighborhood group of jellyfish i, with A and B being random, non-equal integers within the range $[1,\ \mathrm{N_{JF}}]$, and AZi indicating the size of SNG_i. Subsequently, the average position (A_{vi} (- SNG)) of each jellyfish is calculated using data from its respective social neighborhood group. As a result, the average position of all jellyfish, previously calculated by Eq. (10), is replaced by:

$$JF_i(It+1) = Z \times (JF_{\text{Best}} - 3 \times Z \times AV_i(SNG)) + JF_i(It) \tag{16}$$

3.7 Data Normalization

Normalization of log data is crucial in eliminating systematic errors introduced by disparate measurement equipment, enhancing geological data representation, and ensuring uniformity across all log data. Data normalization expedites the training process and mitigates the occurrence of explosive gradient issues. In this study, the MinMaxScaler normalization technique is employed to scale the logging data within the range [0, 1], facilitating consistency and comparability in subsequent analyses.

4 Results and Discussion

The efficacy of the proposed methodology was evaluated using three medical datasets: the Heart Statlog, PIMA Indians Diabetes, and EEG Eyestate datasets. The Heart Statlog dataset encompasses 270 instances with a total of 13 attributes. The PIMA Indians Diabetes dataset contains 768 instances, each characterized by 8 attributes. Finally, the EEG Eyestate dataset comprises 14,980 instances, along with 15 attributes. Table 1 delineates the comprehensive dataset specifications [24–26].

Table 1. Dataset characteristics

Description	Heart Statlog	Pima Indian Diabetes	EEG Eyestate
Total instances	270	768	14,980
Number of attributes	13	8	15
Number of classes	2	2	2
Instances in Class 1	150	268	85,527
Instances in Class 2	120	500	6,723

4.1 Performance Evaluation

The performance of the proposed methodology was assessed using various metrics such as accuracy, sensitivity, specificity, and the kappa index. These metrics provide reliable information about the effectiveness of the approach being tested. Sensitivity, specificity, accuracy, and kappa index for retinal blood detection were calculated as per the general formulae given in Eqs. (17)-(20):

$$\textit{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{17}$$

$$\textit{Specificity} = \frac{TN}{TN + FP} \times 100 \tag{18}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{19}$$

$$Kappa\ index = \frac{Accuracy - Accuracy_T}{1 - Accuracy_T} \tag{20}$$

Table 2 presents a comparative analysis of the proposed classifier against two existing methods across three datasets, employing the aforementioned metrics.

Table 2. Comparative analysis of the proposed classifier with existing methods on binary classification

Datasets	Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)	Kappa Index (%)
Heart Statlog	DBN	72.33	76.55	72.03	86
	CNN	86.95	83	87.33	79.86
	GRU	91.77	88.4	92	85.45
	DCGRU	**97.34**	**97.49**	**96.89**	**88**
Pima Indian Diabetics	DBN	85.43	75	91.22	79
	CNN	89	83.50	90.89	79.08
	GRU	94.76	87	93.98	83.44
	DCGRU	**97.12**	**98.09**	**97.43**	**89.67**
EEG Eyestates	DBN	89.10	92.16	91.40	90.16
	CNN	93.17	93.14	92.71	92.14
	GRU	96.14	93.87	94.36	95.03
	DCGRU	**98.07**	**95.40**	**96.17**	**97.16**

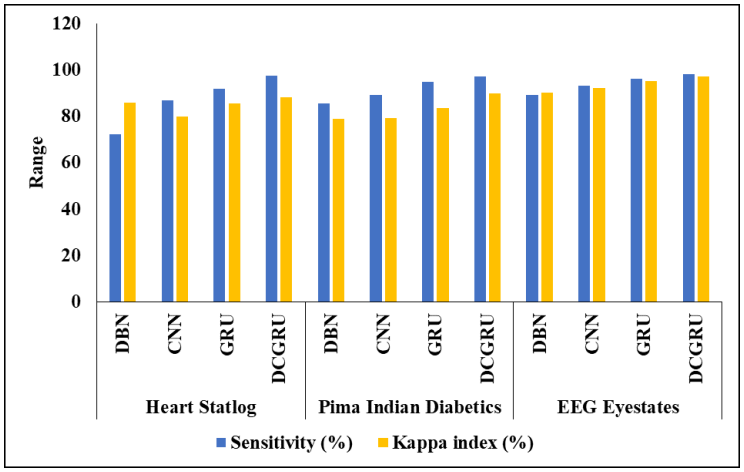

Figure 3. Graphical analyses of various models

In the Heart Statlog dataset, the DBN model exhibited a sensitivity of 72.33%, specificity of 76.55%, accuracy of 72.03%, and a kappa index of 86%. The CNN model achieved a sensitivity of 86.95%, specificity of 83%, accuracy of 87.33%, and a kappa index of 79.86%. The GRU model showed a sensitivity of 91.77%, specificity of 88.4%, accuracy of 92%, and a kappa index of 85.45%. The DCGRU model outperformed with a sensitivity of 97.34%, specificity of 97.49%, accuracy of 96.89%, and a kappa index of 88%. In the Pima Indian Diabetes dataset, the DBN model's sensitivity was 85.43%, specificity 75%, accuracy 91.22%, and kappa index 79%. The CNN model demonstrated a sensitivity of 89%, specificity of 83.50%, accuracy of 90.89%, and kappa index of 79.08%. The GRU model indicated a sensitivity of 94.76%, specificity of 87%, accuracy of 93.98%, and kappa index of 83.44%. The DCGRU model achieved a sensitivity of 97.12%, specificity of 98.09%, accuracy of 97.43%, and kappa index of 89.67%. In the EEG Eyestates dataset, the DBN model recorded a sensitivity of 89.10%, specificity of 92.16%, accuracy of 91.40%, and a kappa index of 90.16%. The CNN model had a sensitivity of 93.17%, specificity of 93.14%, accuracy of 92.71%, and kappa index of 92.14%. The GRU model achieved a sensitivity of 96.14%, specificity of 93.87%, accuracy of 94.36%, and a kappa index of 95.03%. The DCGRU model showed the highest performance with a sensitivity of 98.07%, specificity of 95.40%, accuracy of 96.17%, and a kappa index of 97.16%.

Figures 3 and 4 provide graphical analyses and representations of various models on the three datasets, respectively.

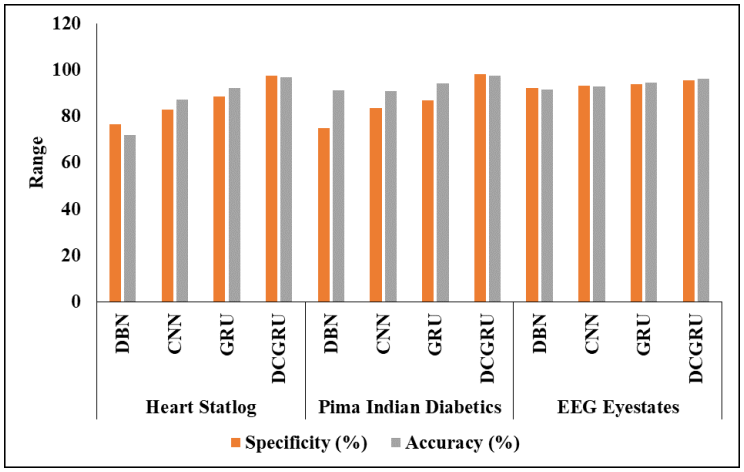

Figure 4. Representations of different models on three datasets

4.2 Evaluation of Proposed Encryption Techniques

Table 3 presents a multi-level analysis of the encryption and decryption time using the proposed encryption method across various data sizes. For the dataset of 20kb, encryption and decryption time were recorded at 0.25 seconds and 0.17 seconds, respectively. At 65kb, these durations were 0.92 seconds for encryption and 0.78 seconds for decryption. Similarly, for the 40kb dataset, encryption and decryption time were observed to be 0.41 seconds and 0.31 seconds, respectively. In the case of the 45kb dataset, the encryption time was 0.58 seconds, and the decryption time was 0.40 seconds. For the 30kb dataset, encryption and decryption time were recorded at 0.4 seconds and 0.30 seconds, respectively.

Figure 5 illustrates a timing analysis of the proposed model across various data sizes.

Table 3. Multi-level encryption period analysis

Size (kb)	Decryption Time (Seconds)	Encryption Period (Seconds)
20 kb	0.17	0.25
65 kb	0.78	0.92
40 kb	0.31	0.41
45 kb	0.40	0.58
30 kb	0.30	0.4

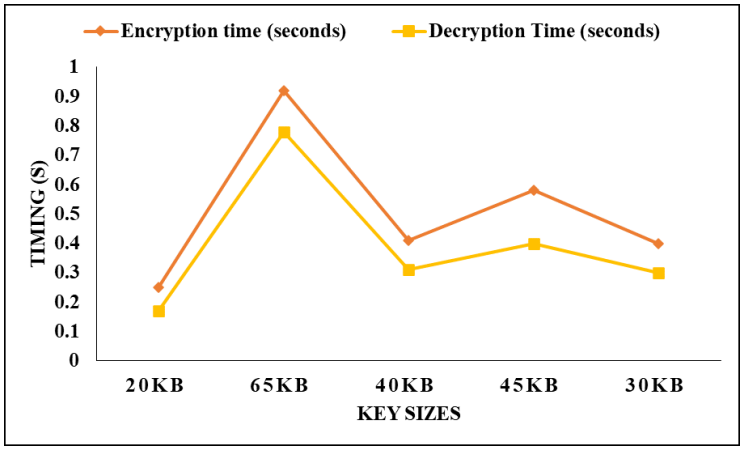

Figure 5. Timing analysis of the proposed model on various data sizes

Table 4 delineates the comparative evaluation of encryption and decryption time across various data sizes for the AES, ECC, and the proposed encryption models. For datasets of 20kb, the AES model exhibited an encryption time of 0.25 seconds and decryption time of 0.60 seconds, whereas the ECC model recorded 0.60 seconds for encryption and 0.79 seconds for decryption. The proposed model demonstrated an encryption time of 0.42 seconds

and a decryption time of 0.70 seconds for the same data size. In the case of the 65kb dataset, the AES model's encryption and decryption time were observed at 0.90 seconds and 1.9 seconds, respectively. The ECC model showed encryption and decryption time of 1.9 seconds and 0.82 seconds, respectively. Conversely, the proposed model achieved encryption and decryption time of 1.60 seconds and 0.72 seconds, respectively. For the 40kb dataset, the AES model's encryption time was 0.40 seconds, and the decryption time was 0.82 seconds. The ECC model exhibited encryption and decryption time of 0.82 seconds and 1.12 seconds, respectively. The proposed model demonstrated encryption and decryption time of 0.72 seconds and 0.98 seconds, respectively. Similarly, for the 45kb dataset, the AES model recorded encryption and decryption time of 0.55 seconds and 1.12 seconds, respectively. The ECC model exhibited encryption and decryption time of 1.12 seconds and 1.9 seconds, respectively. The proposed model's encryption and decryption time were noted at 0.98 seconds and 1.60 seconds, respectively. The average encryption and decryption time for the AES model were 0.5 seconds and 1.046 seconds, respectively, while those for the ECC model were 1.046 seconds and 1.9 seconds. The proposed model achieved an average encryption time of 0.884 seconds and a decryption time of 0.72 seconds. Additionally, the throughput values for the AES, ECC, and the proposed models were 80, 38.2, and 45.42, respectively, for an average data size of 40kb.

Table 4. Comparative analysis of encryption and decryption time

Size (kb)	Time (seconds)		
	AES	ECC	The Proposed Model
20 kb	0.25	0.6	0.42
Average time	0.5	1.046	0.884
Average scope	40	40	40
Throughput	80	38.2	45.42
45 kb	0.55	1.12	0.98
65 kb	0.90	1.9	1.60
40 kb	0.41	0.823	0.72
30 kb	0.50	0.79	0.70

Table 5. Ratio of the encrypted text

Cipher Size to Plain Text Ratio	Encryption Algorithm
5: 2	AES
4: 1	ECC
3: 1	Proposed encryption

Table 5 focuses on the ratio of encrypted text size compared to the plain text size for the AES, ECC, and proposed encryption models. In the AES model, the cipher size to plain text ratio was 5:2. In the ECC model, the ratio was 4:1, and for the proposed encryption model, it was 3:1.

5 Conclusion

Contemporary research focusing on the security of extensive datasets has intensified, acknowledging the significant role various factors play in safeguarding big data. The prioritization of these factors is crucial in the advancement of big data security. Central to this study is the implementation of secure data storage using blockchain technology, augmented by robust user authentication and a health status prediction mechanism. The proposed system integrates an array of technologies, including health condition monitoring, data encryption, and blockchain applications. Patient medical records are securely maintained through blockchain technology. The modified JSO algorithm is employed for feature selection, facilitating the development of a health status monitoring system that predicts patient health using DCGRU. Additionally, the twofish algorithm is suggested as a means of securely transmitting patient IoT data within the computational system.

This study's experimental validation encompassed accuracy, security, and encryption and decryption efficiency. Comparisons with standard methodologies indicate that the proposed approach surpasses existing best practices in effectiveness. Future research could explore the implementation of hyperparameter optimizers and learning rate schedules for the DCGRU model using the described methodology. One potential challenge in employing encryption algorithms like twofish is the computational overhead, especially in handling large datasets. This aspect could affect the real-time processing capabilities of IoT devices, possibly resulting in delays in data transmission and analysis.

Furthermore, there is a growing need to address the ethical considerations surrounding the application of emerging technologies in healthcare. This includes navigating issues related to patient consent, ensuring transparency in

technological applications, and equitable access to advanced healthcare solutions. It is imperative that these ethical dimensions are thoroughly considered to maintain the integrity and trustworthiness of healthcare advancements.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] P. G. Shynu, V. G. Menon, R. L. Kumar, S. Kadry, and Y. Nam, "Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing," *IEEE Access*, vol. 9, pp. 45 706–45 720, 2021. https://doi.org/10.1109/ACCESS.2021.3065440

[2] M. Wang, H. Zhang, H. Wu, G. Li, and K. Gai, "Blockchain-based secure medical data management and disease prediction," in *Proceedings of the Fourth ACM International Symposium on Blockchain and Secure Critical Infrastructure*, 2022, pp. 71–82. https://doi.org/10.1145/3494106.3528678

[3] A. Behrouz and M. Seltzer, "Anomaly detection in multiplex dynamic networks: From blockchain security to brain disease prediction," *arXiv preprint*, p. arXiv:2211.08378, 2022. https://doi.org/10.48550/arXiv.2211.08378

[4] A. Nouman and S. Muneer, "A systematic literature review on heart disease prediction using blockchain and machine learning techniques," *Int. J. Comput. Innov. Sci.*, vol. 1, no. 4, pp. 1–6, 2022.

[5] H. Hasanova, M. Tufail, U. J. Baek, J. T. Park, and M. S. Kim, "A novel blockchain-enabled heart disease prediction mechanism using machine learning," *Comput. Electr. Eng.*, vol. 101, p. 108086, 2022. https://doi.org/10.1016/j.compeleceng.2022.108086

[6] K. Azbeg, O. Ouchetto, S. J. Andaloussi, and L. Fetjah, "A taxonomic review of the use of IoT and blockchain in healthcare applications," *IRBM*, vol. 43, no. 5, pp. 511–519, 2022. https://doi.org/10.1016/j.irbm.2021.05.003

[7] S. Sai, V. Chamola, K. K. R. Choo, B. Sikdar, and J. J. Rodrigues, "Confluence of blockchain and artificial intelligence technologies for secure and scalable healthcare solutions: A review," *IEEE Internet of Things J.*, 2022. https://doi.org/10.1109/JIOT.2022.3232793

[8] M. U. Nasir, M. Zubair, T. M. Ghazal, M. F. Khan, M. Ahmad, A. U. Rahman, H. A. Hamadi, M. A. Khan, and W. Mansoor, "Kidney cancer prediction empowered with blockchain security using transfer learning," *Sensors*, vol. 22, no. 19, p. 7483, 2022. https://doi.org/10.3390/s22197483

[9] M. Chen, T. Malook, A. U. Rehman, Y. Muhammad, M. D. Alshehri, A. Akbar, M. Bilal, and M. A. Khan, "Blockchain-enabled healthcare system for detection of diabetes," *J. Inf. Secur. Appl.*, vol. 58, p. 102771, 2021. https://doi.org/10.1016/j.jisa.2021.102771

[10] A. Rehman, S. Abbas, M. A. Khan, T. M. Ghazal, K. M. Adnan, and A. Mosavi, "A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique," *Comput. Biol. Med.*, vol. 150, p. 106019, 2022. https://doi.org/10.1016/j.compbiomed.2022.106019

[11] S. M. Nagarajan, P. Anandhan, V. Muthukumaran, K. Uma, and U. Kumaran, "Security framework for IoT and deep belief network-based healthcare system using blockchain technology," *Int. J. Electron. Bus.*, vol. 17, no. 3, pp. 226–243, 2022. https://doi.org/10.1504/IJEB.2022.124324

[12] L. Ismail, A. Hennebelle, H. Materwala, J. A. Kaabi, P. Ranjan, and R. Janardhanan, "Secure and privacy-preserving automated end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction," *arXiv preprint*, p. arXiv:2211.07643, 2022. https://doi.org/10.48550/arXiv.2211.07643

[13] S. Saif, S. Biswas, and S. Chattopadhyay, "Intelligent, secure big health data management using deep learning and blockchain technology: An overview," *Deep Learn. Tech. Biomed. Health Inform.*, pp. 187–209, 2020. https://www.doi.org/10.1109/JIOT.2020.3045653

[14] V. Hassija, S. Batra, V. Chamola, T. Anand, P. Goyal, N. Goyal, and M. Guizani, "A blockchain and deep neural networks-based secure framework for enhanced crop protection," *Ad Hoc Netw.*, vol. 119, p. 102537, 2021. https://doi.org/10.1016/j.adhoc.2021.102537

[15] D. Polap, G. Srivastava, A. Jolfaei, and R. M. Parizi, "Blockchain technology and neural networks for the internet of medical things," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Toronto, Canada, 2020, pp. 508–513. https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162735

[16] S. Selvarajan, G. Srivastava, A. O. Khadidos, A. O. Khadidos, M. Baza, A. Alshehri, and J. C. W. Lin, "An artificial intelligence lightweight blockchain security model for security and privacy in IIoT systems," *J. Cloud Comput.*, vol. 12, no. 1, p. 38, 2023. https://doi.org/10.1186/s13677-023-00412-y

[17] F. Kamalov, M. Gheisari, Y. Liu, M. R. Feylizadeh, and S. Moussa, "Critical controlling for the network security and privacy based on blockchain technology: A fuzzy dematel approach," *Sustainability*, vol. 15, no. 13, p. 10068, 2023. https://doi.org/10.3390/su151310068

[18] M. N. Alruwaill, S. P. Mohanty, and E. Kougianos, "hChain: Blockchain based healthcare data sharing with enhanced security and privacy location-based-authentication," in *Proceedings of the Great Lakes Symposium on VLSI 2023*, 2023, pp. 97–102. https://doi.org/10.1145/3583781.3590255

[19] X. Li, H. Zhao, and W. Deng, "BFOD: Blockchain-based privacy protection and security sharing scheme of flight operation data," *IEEE Internet Things J.*, 2023. https://doi.org/10.1109/JIOT.2023.3296460

[20] S. Fang, Q. Liu, F. Zhang, N. Chen, and X. Li, "Application of Internet of Things and blockchain in information security and privacy protection of global organizations," *J. Organ. End User Comput.*, vol. 35, no. 3, pp. 1–16, 2023. https://doi.org/10.4018/JOEUC.323192

[21] A. Ali, B. A. S. Al-Rimy, F. S. Alsubaei, A. A. Almazroi, and A. A. Almazroi, "Healthlock: Blockchain-based privacy preservation using homomorphic encryption in Internet of Things healthcare applications," *Sensors*, vol. 23, no. 15, p. 6762, 2023. https://doi.org/10.3390/s23156762

[22] N. K. Dewangan, P. Chandrakar, S. Kumari, and J. J. Rodrigues, "Enhanced privacy-preserving in student certificate management in blockchain and interplanetary file system," *Multimed. Tools Appl.*, vol. 82, no. 8, pp. 12 595–12 614, 2023. https://doi.org/10.1007/s11042-022-13915-8

[23] B. Schneier, J. Kelsey, D. Whiting, D. Wagner, C. Hall, and N. Ferguson, "Twofish: A 128-bit block cipher," *AES Submiss.*, vol. 15, pp. 23–91, 1998.

[24] Statlog (Heart), "UC Irvine machine learning repository." http://archive.ics.uci.edu/ml/datasets/statlog+(heart).

[25] Pima Indians Diabetes Database, "UCI machine learning." https://www.kaggle.com/uciml/pima-indians-diabetes-database.

[26] EEG Eye State, "UC Irvine machine learning repository." https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State.