Privacy-Enhanced Heart Disease Prediction in Cloud-Based Healthcare Systems: A Deep Learning Approach with Blockchain-Based Transmission

Ahmad Raza Khan[1,*], Abdul Khader Jilani[2]
[1] Department of Information Technology, College of Computer and Information Sciences, Majmaah University, AlMajmaah, 11952, Saudi Arabia
[2] Department of Computer Science, University of Technology Bahrain, Bahrain
Emails: ar.khan@mu.edu.sa; a.jilani@utb.edu.bh

Abstract

The increasing adoption of cloud computing in healthcare presents immense opportunities for disease prediction, while raising critical privacy concerns. This study proposes a novel privacy-preserving scheme that leverages advanced cryptographic techniques, blockchain technology and deep learning approach within a cloud platform, to ensure secure data handling and accurate disease prediction. The proposed methodology encompasses authentication, encryption, blockchain-based transmission, and a deep learning-based heart disease prediction system (HDPS). Through rigorous authentication protocols and two-level security mechanisms, patient data is securely encrypted using RSA and Blowfish encryption before storage in the cloud. Blockchain technology facilitates secure data transmission, ensuring integrity and traceability. At the receiver end, data decryption precedes input into the HDPS, comprising artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNN). The HDPS incorporates data preprocessing, feature extraction, feature selection, and a deep learning-based prediction model, achieving remarkable accuracy (0.9941) in heart disease prediction. Implemented in MATLAB, this approach offers a robust framework for privacy-preserving heart disease prediction in cloud-based healthcare systems.

Keywords: Cloud Computing; Privacy-Preserving Scheme; Heart Disease Prediction; Blockchain-Based Transmission; Two-Level Security Mechanism

1. Introduction

Heart disease prediction is estimating a person's chance of getting cardiovascular problems by captivating into interpretation their age, gender, routine, and health, among other criteria [1]. Cardiovascular diseases (CVD) are presently the leading cause of death universally, with a projected annual death toll of around 17.9 million according to the World Health Organization's 2020 report [2]. The WHO has estimated that cardiovascular disease sources about 17.9 million deaths annually globally. Among this loss of life, 80% are attributed to coronary artery disease and cerebral stroke [3]. Heart disease prediction is an essential element of preventive care techniques in contemporary healthcare systems, to diminish the impact of cardiovascular illnesses. By harnessing technological breakthroughs and data analytics, healthcare practitioners employ several methodologies to forecast the probability of cardiac disease in patients. These methods frequently entail the usage of electronic health records (EHRs), wearable devices, and machine learning algorithms [4].

Predictive models can identify persons at increased risk of developing heart disease by assessing comprehensive patient data, encompassing demographics, medical history, lifestyle variables, and physiological indicators such as blood pressure and cholesterol levels. The use of cloud computing technology in contemporary healthcare systems has considerable opportunities for enhancing cardiac disease prediction but is accompanied by apprehensions surrounding data confidentiality and safeguarding [5]. Ensuring the security of patients' sensitive medical information is crucial in cloud-based healthcare systems (HCS), requiring the implementation of a privacy-preserving approach for heart disease prediction. Cloud computing in healthcare refers to the practice of storing, organizing, and analyzing medical data on remote servers that may be accessed over the Internet [6]. This strategy provides the potential to easily expand, adapt, and achieve cost-efficiency, allowing healthcare practitioners to effectively manage substantial

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

amounts of data for predictive analytics and other uses. Nevertheless, the concentrated structure of cloud infrastructure presents concerns about privacy and security, specifically with the safeguarding of individuals' personal health information (PHI) [7].

To reduce these potential dangers, a privacy-preserving strategy is used in cloud-based healthcare systems to anticipate cardiac disease. Homomorphic encryption is a method used to maintain privacy in cloud-based healthcare systems [8]. It allows controls to be carried out on encrypted data without the essential for decryption. Through the process of encrypting patients' medical records before their transfer to the cloud, healthcare providers may do predictive analytics on the encrypted data in a safe manner, therefore preventing illegal access to critical information [9]. Differential privacy is a further method used to safeguard individual privacy while maintaining accurate aggregate analysis by introducing noise into query results. Differential privacy guarantees that the result of a query does not disclose any confidential details about particular patients, therefore protecting their privacy when it comes to predicting heart disease [10]. Secure multiparty computing (SMC) enables healthcare providers to engage in collaborative predictive analytics for heart disease prediction while ensuring privacy by not disclosing patients' raw data. This approach allows healthcare providers to leverage collective intelligence without compromising privacy.

Access control systems, such as role-based access control (RBAC) and attribute-based access control (ABAC), impose limitations on the individuals authorized to access patients' medical information and the specific conditions under which they can do so [11]. Healthcare organizations may safeguard the confidentiality of sensitive information related to heart disease prediction by establishing clear roles, permissions, and data access regulations, therefore restricting access to authorized persons only. In addition, robust data anonymization approaches, such as k-anonymity and l-diversity, safeguard the identity of patients by removing their personal information from medical records before storing or distributing them in the cloud [12]. Anonymization techniques safeguard against re-identification of people by eliminating or simplifying identifying information, such as names and social security numbers. This allows for meaningful research on heart disease risk variables [13]. A privacy-preserving scheme for heart disease prediction in modern HCS utilizes cryptographic techniques, privacy-enhancing technologies, access control mechanisms, and blockchain technology to protect patients' privacy while facilitating precise and efficient predictive analytics. This method guarantees the security, integrity, and availability of patient's sensitive medical information in cloud-based healthcare systems, effectively resolving privacy issues.

Key contributions:

- **Dual-layer encryption:** Data is protected using a robust combination of RSA and Blowfish algorithms, safeguarding both transmission and storage.
- **Blockchain-based transmission:** Leveraging Ethereum's tamper-proof platform enhances data integrity and traceability.
- **Optimized Deep Learning model:** A hybrid architecture incorporating ANN, CNN, and RvNN achieves high prediction accuracy (0.9941), surpassing single-model approaches.
- **Improved feature selection:** The Enhanced Information Gain method ensures efficient feature selection, reducing computational overhead and model complexity.

The rest of this paper is arranged as: Section 2 provides a clear review on the state-of-art approaches that has been the benchmark for the study. Section 3 provides a clear explanation on the proposed approach, and Section 4 provides the interpretation of the acquired outcome's This article is concluded with a future scope in Section 5.

2. Literature Review

Padinjappurathu *et al.*, (2022) [14] presented an effective PP approach for patient healthcare data obtained from IoT strategies for illness calculation in the current HCS. After authentication, the suggested system uses log-of-round value-based Elliptic Curve Cryptography (LR-ECC) to secure data transport. Authorized hospital workers can safely download patient data. The EHGA-DLNN may evaluate this data with the trained system to forecast illnesses. The investigational results show that the suggested method increases prediction accuracy, privacy, and security over previous approaches.

Jayaram and Prabakaran (2021) [15] Provided edge-layer additive homomorphic encryption that protects privacy while handling data safely and removing insensitive information. In the suggested HCS, edge-level screening and offload reduce response times and network bandwidth. For the purpose of distant illness prediction and rehabilitation, put forth a cloud-layer adaptive weighted probabilistic classifier algorithm. Compared to classifiers, it will increase

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

the rapidity and precision of illness prediction. Lastly, utilizing data on Parkinson's disease, the SECHS was evaluated for both reliability and safety.

Munirathinam *et al.*, (2020) [16] described a unique e-healthcare system for deceased sickness monitoring using IoT, Cloud, deep learning, fuzzy rules, and temporal features. Patients utilizing e-healthcare devices send medical data to this system. A revolutionary cloud storage mechanism secures recovered and encrypted data. Second, decryption restores data. Third, a unique cloud architecture predicts heart rate and diabetes using UCI Repository dataset medical data. A revolutionary deep learning algorithm predicts sickness severity using the Convolutional Neural Network.

Two reliable and privacy-preserving Top-k disease matching methods are suggested by Xu *et al.*, (2020) [17]. The first option generates k diagnostic files using our weighted Euclidean distance comparison algorithm and safe k-nearest neighbor technique. Weight each body signal as needed. The second method compares Euclidean distances using a superlinear sequence and the modified Paillier homomorphic encryption algorithm to reduce computational and transmission overhead. The trustworthy party does not need encryption, but the user side has a much higher computing overhead. The two systems can be employed in different scenarios. Security analysis and synthetic and real data simulations show the schemes' efficiency and privacy-preserving properties.

Cloud computing and SVM are utilized to forecast cardiac problems by Khan *et al.*, (2020) [18]. Simulations reveal that the intelligent cloud-based heart disease prediction system with a SVM-based system model has 93.33% accuracy, better than earlier techniques. Health facility technology innovation is helping manage patients with various ailments. Heart disease, which cannot be seen, is the deadliest and strikes when vital indicators like pulse rate, physique infection, and blood pressure surpass the normal range. The actual trouble is diagnosing patients more accurately and quickly, then prescribing suitable therapies and minimizing prescription mistakes.

Nancy *et al.*¸ (2020) [19] combined with the Cloud, the IoT connects people and things, improving our lives. As AI and ML become more widespread in healthcare, predictive analytics can help convert reactive strategies into proactive ones. Deep learning, a subset of ML, can analyze massive data sets quickly, provide intelligent insights, and solve complex problems. Preventive treatment and early intervention for at-risk individuals require accurate and rapid illness prediction. With the increased deployment of electronic medical archives, accurate prediction methods are important to using RNN types of DL that can manage sequential time-series data. IoT devices provide data, and cloud-stored healthcare data on patient history is predictively analyzed.

Li *et al.*¸ (2020) [20] presented a private-preserving, multi-party computation-based, secure self-service diagnosis technique. Once the patient encrypts and transmits their health data, the medical facility's server calculates the level of similarity among the patient's data and the trait vector of the clinic illness. At last, the healthcare system uses the calculated resemblance value to identify the patient's ailment and provides the appropriate course of therapy. The self-serviced medical diagnostic scheme uses HE and privacy-preserving access management to secure patient physical heath data and facility diagnostic modal confidentiality. Provide thorough safety evaluation to show that the technique is resilient to recognized safety hazards.

Wang *et al.*¸ (2023) [21] suggested FRESH, a broad smart healthcare system for exchanging physiological data, using FL and ring signature defense from threats. Wearable gadgets capture physiological data in FRESH. Edge computers (e.g., phones, and tablets) train ML models using local data. To train FL illness prediction models, edge computing devices submit model parameters to the central server. The source of parameter changes during cooperative FL training is hidden via certificate ring signature to combat SIAs. An improved batch verification process in the suggested ring signature schema uses the additivity of linear operations on elliptic curves to minimize server processing effort.

Sharma *et al.*, (2018) [22] discussed the technical challenges of integrating practical privacy-preserving algorithms into medical records. Utilizing contemporary algorithms and gadgets, such as IoT gadgets and clouds, current healthcare systems gather and evaluate individual medical information at previously unheard-of depths and sizes. Models of analysis derived from these kinds of data sources are used by physicians, investigators, patients, and healthcare workers to diagnose illnesses, create customized medication regimens, and track patients. The research makes use of kHealth, a personalized electronic health information system that is being created and evaluated for disease tracking.

Chenthara *et al.,* (2019) [23] discussed cyber security research issues and directions. Several publications were studied, examined, and evaluated to determine the following tasks: 1) EHR security and privacy; 2) cloud e-health data security and privacy; 3) cloud architecture; and 4) different cryptographic and non-cryptographic EHR techniques. Big data provides a wealth of information and expertise for e-health applications, creating urgent privacy and security issues.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

2.1 Problem Statement

Technology improvements haven't solved the challenge of maintaining privacy and security in healthcare systems, especially when it comes to the diagnosis and treatment of diseases. Sharma *et al.*, (2018) [22] explained the prevalence of personal health data gathering and analysis has increased due to the widespread use of IoT devices, cloud computing, and machine learning in healthcare (Khan *et al.*, 2020) [18]. Nevertheless, the task of safeguarding patients' confidential medical data while enabling efficient predictive analytics continues to pose a difficulty (*et al.*, 2020) [19]. Current methodologies frequently lack adequate privacy-preserving safeguards (Padinjappurathu *et al.*, 2022) [14], hence exposing patient data to illegal access and exploitation. Moreover, the incorporation of many technologies and data sources adds to the complexity of privacy (Munirathinam *et al.*, 2020) [16] and security concerns. To talk about these difficulties, it is necessary to create extensive privacy-preserving solutions that can protect patient data at every stage, including collection, storage, analysis, and sharing (Li *et al.*, 2020) [11]. To secure the security, integrity, and availability of healthcare data, it is crucial to prioritize the implementation of strong encryption, access control, and anonymization techniques (Xu *et al.*, 2020) [17]. These measures will also facilitate accurate and prompt sickness prediction and management.

3. Methodology

Proposed Methodology: Securing Heart Disease Prediction in the Cloud

Our scheme safeguards user privacy while enabling accurate heart disease prediction using cloud computing. It follows a step-by-step approach:

1. Secure Entry:

- **Registration:** Patients register by providing details and medical data (anonymized where possible). Text data from various sources is combined and encrypted using Blowfish with a unique, securely generated key.
- **Login:** Authorized access is ensured through verified credentials and potentially multi-factor authentication.
- **Verification:** Patient identity and data ownership are confirmed, along with access permissions and data usage control.

2. Double-Layer Encryption:

- Data confidentiality is ensured through a two-level mechanism:
    - **Level 1:** Symmetric Blowfish encryption using the patient's key secures their data.
    - **Level 2:** The encrypted data and key are further protected with asymmetric RSA encryption using the cloud server's public key.

3. Secure Storage on the Cloud:

- Encrypted data resides securely on the cloud platform with access control mechanisms to prevent unauthorized access or modifications.

4. Blockchain-Enabled Transmission:

- Data integrity and traceability are enhanced by transmitting it through the Ethereum blockchain network.
- Smart contracts can be implemented for automated access control and data usage agreements.

5. Decryption at the Receiving End:

- The authorized receiver uses the cloud server's private key to decrypt the Level 2 encryption.
- The patient's key retrieved from Level 2 is then used to decrypt the Level 1 Blowfish encryption.

6. Heart Disease Prediction System (HDPS):

- **Pre-processing:** Data is cleaned, missing values are handled, and outliers are identified and addressed. Normalization ensures all data falls within a common range.
- **Feature Extraction:** Relevant features are extracted, capturing both central tendency (mean, median, mode) and degree of dispersion (range, variance, standard deviation).

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

- **Feature Selection:** The Improved Information Gain technique efficiently selects the most informative features for model building.
- **Deep Learning Model:** A hybrid architecture combining ANN, CNN, and RvNN models learns complex patterns from the selected features to predict heart disease probability.

3.1 Double-Layer Secure Entry and Encryption for Patient Data in Healthcare Systems

The proposed methodology comprises two core components to ensure the utmost security and confidentiality of patient data within healthcare systems. Firstly, the "Secure Entry" protocol establishes a robust registration process where patients provide their details and medical data, which are encrypted using Blowfish with a unique key. Upon login, access is granted only to authorized individuals through verified credentials and potentially multi-factor authentication, followed by thorough verification to confirm patient identity, data ownership, access permissions, and data usage control. Secondly, the "Double-Layer Encryption" mechanism reinforces data confidentiality through two distinct levels. At Level 1, symmetric Blowfish encryption safeguards patient data using their dedicated key. Subsequently, at Level 2, the encrypted data and key undergo additional protection via asymmetric RSA encryption, utilizing the cloud server's public key. This comprehensive approach ensures the integrity and confidentiality of patient data at both the entry and encryption stages within healthcare systems.

3.1.1 Components

The four components of the suggested system are monitoring, disease prediction system, secure data transfer, and authentication. Following the installation of the IoT sensor devices on the patient's body, the patient should record with the relevant clinic via the hospital's website or cell phone application. Following a successful login using one of the suggested authentication methods, the sensor readings are detected and securely transferred via the Fog layer into the HCS. Concurrently, the corresponding physician at the hospital may securely transfer the patient's data and compare it with a previously trained model. The overall architecture of the proposed methodology can be shown in Figure 1.

- **Authentication Phase:** To enhance the safety of the model and data transfer, authentication is established between physicians, medical staff, patients, and the Cloud Server (CS), as well as between the healthcare facility and the CS. The suggested system's initial step is this one. Providing access to approved IoT sensor devices requires taking this crucial step. There are three steps in the authentication process
- Registration
- Login
- Verification

Registration*:* Before the data on numerous IoT strategies linked to the HCS can be accessed, the officer must provide their consent. The officer provides data to the IoT device for authentication after verification. These are examples of the four sections that make up the registration procedure.

Patient Details: The user provides the patient's data mostly during the registration process. The health assistant enters and saves the patient details, which include the following: Username, Patient Name, Sex, Age, Address, Password, Patient ID, Hospital ID, Doctor Name, and so forth. The patient's information may be expressed numerically as in Eq. (1)

$$\tilde{P}_{pd} = \{\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \dots \dots \tilde{p}_k\} \quad (1)$$

Here, $\tilde{P}_{pd}$ denotes the patient's data, such as name, age, sex, and patient ID, while Pepd denotes the patient facts set.

Combine Text: Once the patient's information has been entered, combine the two sentences into a single text explaining the concatenation process. In this case, the user ID and the corresponding hospital ID are combined into a single text that may be expressed mathematically as in Eq. (2)

$$\vec{T}''_{ct} = \tilde{p}_u \oplus \tilde{p}_h \quad (2)$$

Here, the combined text is indicated by $\vec{T}''_{ct}$, while the user ID and hospital ID are retrieved from $\tilde{P}_{pd}$ are indicated by $\tilde{p}_u$ and $\tilde{p}_h$, respectively.

Ciphering combined text: After the text has been concatenated as before during registration, the ciphering process is carried out to turn the text into ciphertext using the substitution cipher. A substitution cypher is a kind of encryption

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

in which predefined ciphertext is used instead of textual elements. The simplest substitution cyphers are ones where the normal text letter's cyclical shift serves as the cipher letter. Better security is provided by these alternative ciphertext values. Additionally, the basic twenty-six letters, punctuation, digits, and common syllables may all be simply expanded into the alphabet's plaintext. Letter encryption can be stated mathematically as in Eq. (3)

$$\left(\vec{T}''_{ct}\right)_{encrypt} = \left(\vec{T}''_{ct}\right) mod 26 \quad (3)$$

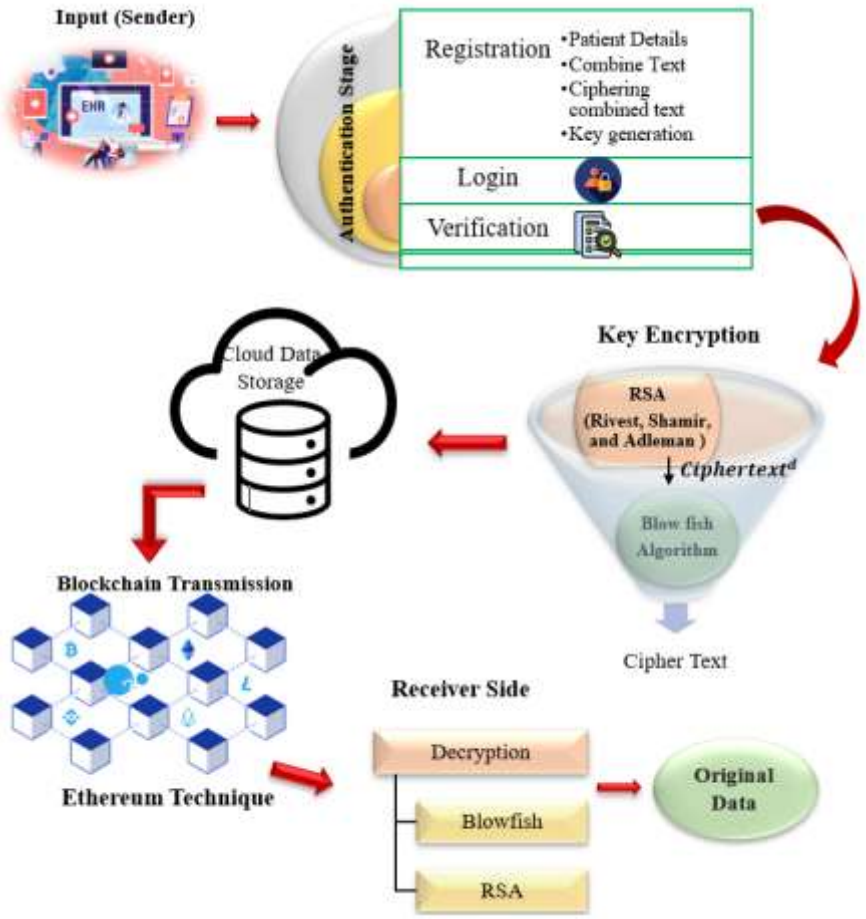

Figure 1: Overall architecture of the proposed methodology

The encrypted text $\left(\vec{T}''_{ct}\right)$ in this case, represents the combined text's ciphertext. The CS sends these ciphertexts to the data owner at the moment of verification. If a user tries to download any file from the CS, the CS asks them to give the ciphertext. The user sends the correct ciphertext, and the CS verifies that the user is allowed and has permission to view the data. The substitution cipher can be expressed mathematically as follows in Eq. (4) and Eq. (5) respectively.

$$\left(\vec{T}''_{ct}\right)_{encrypt} \xrightarrow{Matched} \left(cs \xrightarrow{confirms} \tilde{A}_{au}\right) \quad (4)$$

$$\left(\vec{T}''_{ct}\right)_{encrypt} \xrightarrow{Not\ Matched} \left(cs \xrightarrow{confirms} \tilde{A}_{ur}\right) \quad (5)$$

In this case, $\tilde{A}_{au}$ and $\tilde{A}_{ur}$ stand in for the respective approved and unauthorized users.

Key Generation: The Cloud generates the public and private keys. The public key is supplied, while the private key is emailed to the user after registration, confirming data encryption and key production. The Cloud provider demands the private key upon file viewing requests. When the user provides the right private key, the cloud provider decrypts the text file and displays it to them. When the private key is wrong, the encrypted data format is shown instead of the original file. The mathematical formulation of the public key and cloud-generated private keys is in Eq. (6)

$$CS \xrightarrow{\left(\vec{K}''_{pu}, \vec{K}''_{pr}\right)} User \quad (6)$$

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

The public key is represented by $\vec{K''}_{pu}$, whereas the private key is represented by $\overleftarrow{K''}_{pr}$. The secret key has been calculated to enhance security. To determine the secret key, the round log value of the $\vec{K''}_{pu}$ and $\overleftarrow{K''}_{pr}$ was used, as expressed mathematically in Eq. (7)

$$\vec{K''}_{se} = \log\left(\vec{K''}_{pu} \oplus \overleftarrow{K''}_{pr}\right) \quad (7)$$

Here, $\vec{K''}_{se}$ represents the secret key, while $\oplus$ representing the round log value of the $\vec{K''}_{pu}$ and $\overleftarrow{K''}_{pr}$.

Login: A set of credentials called a login is used to verify a user. They typically include both the password and the username. By providing their username and password, a user can get access to an application through the login section. Patients should provide the login information provided by the administrator when logging into the system. When logging in, the patient has to provide the user's name, password, and ciphertext.

Verification: The system has been logged in once the verification process is completed. The ciphertext, password, user-id, and username for this section would all match in the system. If all the information matches, the algorithm concludes that the patient has already listed with the relevant Hospital Cloud Server. Alternatively, the system goes back to the registration phase.

- **Encryption using a Two-level Security Mechanism**

To expand data protection, the two-level security technique combines the best features of the RSA and Blowfish encryption algorithms. The ciphertext output produced by RSA, which is renowned for its strong asymmetric encryption, is used as the input for Blowfish encryption. This encrypted data from Blowfish is stored in the cloud platform.

RSA

The RSA procedure was the first asymmetric encryption technique, and it was proposed by Rivest, Shamir, and Adleman (RSA) in 1978. A private key, which is distinct from the public key and is only known to the receiver, is used for decryption in the RSA Encryption Algorithm. The product of two huge prime numbers would be the public key. The product is released to the public. There is no known way to determine the prime factors of such numbers, and decryption would need knowledge of the number's two prime factors. This implies that the private key can only be produced by the same person who generated the public key.

Step 1: Select two big prime numbers as $a\ and\ b$
Step 2: Compute $n = a * b$
Step 3: Choose a public key for encryption, such as E. Confirm that this key does not depend on $(a - 1)\ or\ (b - 1)$.
Step 4: Select an integer E such that $1 < E < z$.
Step 5: Compute $(d * e)\ mod\ (a - 1) * (b - 1) = 1$
Step 6: A pair of private keys can be bundled as (n, d) and a pair of public keys can be bundled as (n, e)
Step 7: Calculate ciphertext from plaintext using $Ciphertext = Plaintext^{e} mod\ n$
Step 8: Sending the receiver the ciphertext when it has been produced
Step 9: Decrypt the plaintext from the ciphertext using $Plaintext = Ciphertext^{d} mod\ n$. This $Ciphertext^{d}$ enters as input to the blowfish model.

Blowfish encryption technique

Blowfish is a symmetric encryption technique, a 64-bit block cipher with a maximum key length of 448 bits. Bruce Schneier developed it in 1993 as a replacement for the antiquated International Data Encryption Algorithm (IDEA) and Data Encryption Standard (DES) encryption techniques. Blowfish is renowned for its simplicity and quickness, although its use has recently decreased. Newer, more safe encryption methods like the Advanced Encryption Standard (AES) are replacing it. The Blowfish encryption process can be shown in Figure 2

Expansion of cryptographic keys: The secret key that Blowfish starts with can range in length from 32 to 448 bits. Next, to create many subkeys, the encryption key is produced and extended using the P-array and S-boxes $S[x]$ precomputation.

Creation of subkeys: Two 32-bit chunks are created from the 64-bit blocks that make up the stretched-out key. To create a new collection of subkeys $\{S[0], S[1] \ldots.\}$, these parts are combined with a few specified values. In both the encryption and decryption processes, four substitution boxes $\{S[0], S[1], S[2], S[3]\}$, with 255 entities $\{S[0], S[1] \ldots. S[255]\}$ with 32 bits are required.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

Encryption of data: The major portion is about to begin. These two 32-bit parts are encrypted sixteen times. A difficult series of transpositions and replacements (XOR operations, additions, and lookups in the S-boxes) are involved in each round.

Post-processing: 64-bit ciphertext blocks are created by rejoining the 32-bit jumbled fragments after the 16 rounds.

Decryption: In Blowfish, decryption is accomplished by reversing the encryption process. Consequently, everything reverses until the ciphertext is converted back into plaintext.

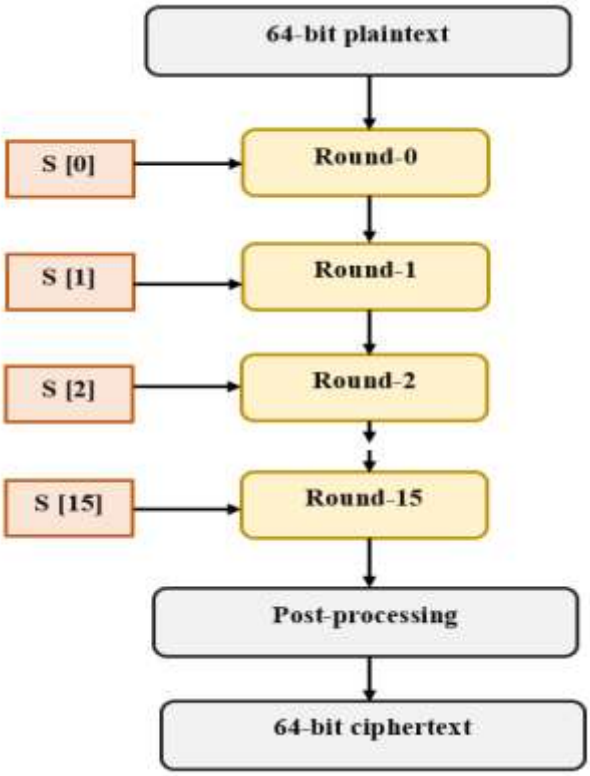

Figure 2: Blowfish Encryption process

3.2 Cloud Data storage

After encryption (by two-level security mechanism), the cipher texts, which are securely encoded representations of the original data, are saved in a cloud storage database. Cloud storage provides flexible and easy options for storing data, enabling encrypted data to be accessed remotely as required. Utilizing encrypted data storage mechanisms, sensitive information can be safeguarded from unauthorized access or modifications. Robust access control mechanisms, such as multi-factor authentication and role-based access control, further fortify the security posture of cloud-stored data. Encryption keys are managed securely, and regular audits are conducted to detect and mitigate potential vulnerabilities. Additionally, data backup and disaster recovery protocols are implemented to ensure business continuity in the event of unforeseen incidents.

Then, transmission takes place by employing the Ethereum technology's blockchain chain transmission. The decentralized ledger of blockchain ensures that every transaction, including the transfer of encrypted data, is recorded and validated by several nodes, ensuring the immutability and transparency of data. The Ethereum protocol strengthens security by offering a robust framework for securely transferring encrypted information, ensuring the privacy and integrity of sensitive data throughout transmission.

3.3 Blockchain transmission using Ethereum Technology

Leveraging blockchain technology, particularly the Ethereum network, enhances data integrity and traceability during transmission processes. By employing distributed ledger technology, each data transaction is recorded in a tamper-resistant and immutable manner, providing a transparent and auditable trail of data activity. Smart contracts, programmable self-executing agreements, can be utilized to automate access control and enforce predefined data usage agreements. These smart contracts enable conditional execution of actions based on predefined criteria, ensuring that data is accessed and utilized in accordance with established protocols. Through the decentralized nature of blockchain

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

networks, data transmission processes are made more resilient to tampering or unauthorized alterations, bolstering trust and confidence in data exchanges.

Ethereum is a decentralized blockchain platform that leads the way in peer-to-peer transactions enabled by smart contracts, which are self-executing agreements recorded in code. Figure 3 shows the Ethereum technology. These contracts essentially provide interactions without the need for trust, bypassing the requirement for a central authority, and enabling users to directly engage in transactions with each other. The blockchain technology of Ethereum ensures that transactions cannot be changed, can be verified, and are decentralized, representing a groundbreaking model in digital commerce. Users interact with Ethereum by utilizing user-generated accounts, each equipped with the capacity to authenticate transactions using Ether, the native digital currency of Ethereum. Ether is used as a fee for executing transactions on the network.

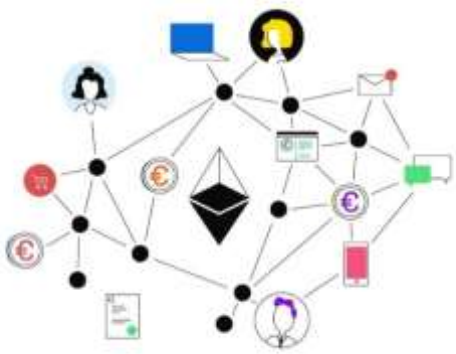

Figure 3: Ethereum Technology

Ether performs an important character in powering the execution of smart contracts and enabling diverse interactions within the Ethereum ecosystem. Smart contracts, coded in the Solidity programming language, have a crucial function in automating the execution of contracts. These contracts define the conditions of the agreement in a clear and executable way by encapsulating functions, events, and state variables. After being constructed, smart contracts are transformed into bytecode and then published onto the Ethereum blockchain. Each contract is given a distinct address, which serves as its unique identifier within the network. The execution of smart contract code takes place within the Ethereum Virtual computer (EVM), which is a Turing complete stack-based virtual computer. The EVM offers a segregated runtime environment that guarantees the integrity and security of contract execution by isolating it from the network. Significantly, the EVM's computational capabilities are limited by transaction fees, which function as a means to prevent misuse and guarantee the effective distribution of network resources. The Ethereum ecosystem places great importance on security, partly because smart contracts become unchangeable once they are published on the blockchain. Consequently, smart contract developers need to follow strict coding practices and take into account any vulnerabilities while developing. Neglecting to comply with this requirement may result in significant repercussions, such as financial losses and compromised network security. Therefore, the Ethereum community prioritizes continuous research, education, and adherence to best practices to cultivate a safe and resilient environment for decentralized apps and smart contracts. Figure 4 demonstrates the blockchain transmission by Ethereum Technology.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

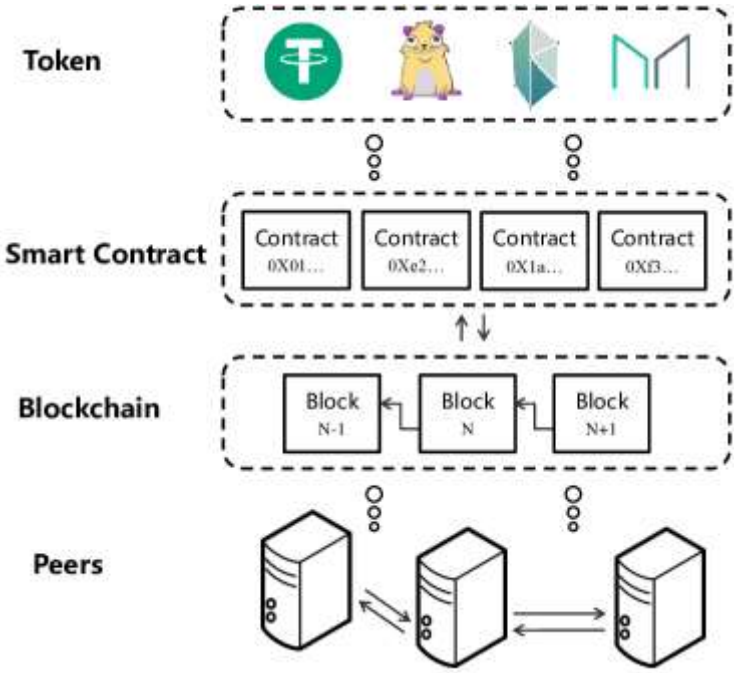

Figure 4: Blockchain transmission using Ethereum Technology

- **Receiving Process**

In secure communication, the decryption procedure requires the receiver to hold the private key that is used to decrypt communications that have been encrypted utilizing the associated public key. After receiving the ciphertext, the receiver employs their private key to decipher the message, guaranteeing both secrecy and integrity. This method is crucial for ensuring the safe transfer of data, with a particular focus on the secure exchange of information in different situations, such as secure messaging, online transactions, and data sharing via networks. While asymmetric encryption utilizes the recipient's private key to reverse the encryption created by the sender's public key, symmetric encryption uses the same key for both encryption and decryption. Moreover, Blowfish decryption employs the same key to undo the encryption process, guaranteeing both effective and safe retrieval of the original plaintext. RSA encryption, derived from the Rivest-Shamir-Adleman algorithm, is essential for ensuring secure communication. It does this by creating pairs of public and private keys using big prime numbers, so offering strong cryptographic protection for the transfer of sensitive information.

3.4 Heart Disease Prediction System (HDPS)

At the receiver end, typically a healthcare professional such as a doctor, the cloud-stored data is decrypted using a two-level security mechanism. Once decrypted, the data undergoes detection for heart disease using the trained deep learning model. This detection process analyzes the extracted features to provide accurate predictions of heart disease probability, enabling healthcare professionals to make informed decisions regarding patient care and treatment strategies. The Heart Disease Prediction System (HDPS) employs a comprehensive approach to accurately predict the likelihood of heart disease in individuals.

- **Pre-processing:** Data is cleaned, missing values are handled, and outliers are identified and addressed. Normalization ensures all data falls within a common range.
- **Feature Extraction:** Relevant features are extracted, capturing both central tendency (mean, median, mode) and degree of dispersion (range, variance, standard deviation).
- **Feature Selection:** The Improved Information Gain technique efficiently selects the most informative features for model building.
- **Deep Learning Model:** A hybrid architecture combining ANN, CNN, and RvNN models learns complex patterns from the selected features to predict heart disease probability.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

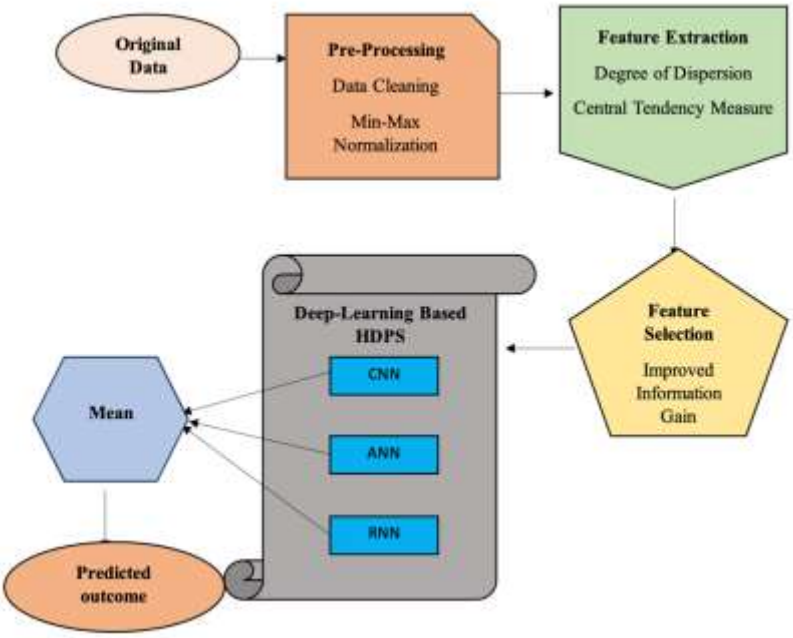

Figure 5: Architecture of HDPS

3.1.1 Data collection

The data are composed from the Hungarian dataset. Cardiovascular diseases (CVDs), commonly referred to as heart disease, are the most common reason of death universal, accounting for 17.9 million deaths annually or almost 32% of all fatalities. Cardiac and vascular conditions collectively known as cardiovascular diseases (CVDs) include stroke, coronary heart disease, and rheumatic heart failure. Four out of every five CVD fatalities are caused by cardiovascular events, with the majority of these premature deaths occurring in those below the age of 70.

3.1.2 Pre-Processing

The gathered original data are pre-processed via Data cleaning (Data Deduplication, Missing value imputation) and MinMax Normalization. Figure 6 shows the pre-processing technique.

Figure 6: Pre-Processing

3.1.2.1 Data Cleaning

Data cleaning involves removing incorrect, corrupted, duplicate, or incomplete data from a dataset, ensuring reliable outcomes and algorithms by addressing potential issues with multiple data sources.

- ***Data Deduplication***

Data deduplication is a technique that lowers storage expenditures by getting rid of duplicate data. A pointer to the unique data copy is used in place of redundant data blocks. Data deduplication and incremental backup are similar in that they copy just the data that has altered since the last backup.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

- ***Missing value imputation***

Missing value imputation refers to the process of filling in or estimating the values that are absent or not available in a dataset. When certain data points are missing for various reasons such as errors, non-responses, or data collection issues, imputation methods are applied to estimate or substitute those missing values. The goal is to maintain the integrity and usefulness of the dataset for analysis or modelling purposes. Different imputation techniques can be used, ranging from modest methods like mean or median imputation to more complex approaches such as regression-based imputation or machine learning algorithms.

3.1.2.2 MinMax Normalization

With this method, the numerical values of a feature are adjusted to range from 0 to 1. To do this, subtract the smallest value of the feature from every value, then divide the resulting number by the entire range of the feature.

$$A_{scaled} = \frac{A - A_{min}}{A_{max} - A_{min}} \quad (8)$$

Where, A_{scaled} is the new value of feature A which is scaled, A-old value of feature A, A_{min} is the minimum value of the feature A, A_{max} is the maximum value of the feature A.

3.1.3 Feature Extraction

From the pre-processed data the features like Central Tendency Measure (Harmonic Mean, Mode, Median) and Degree of Dispersion Measure (Range, Variance, SD) are extracted. Figure 7 shows the Feature Extraction techniques.

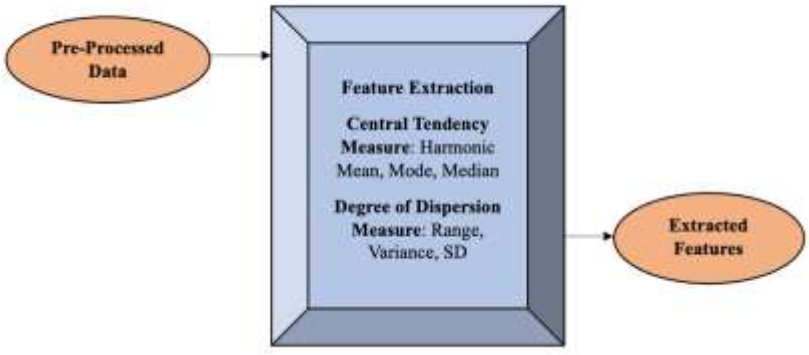

Figure 7: Feature Extraction

3.1.3.1 Central Tendency Measure

The numbers that are employed to indicate the mid-value or the center value of an extensive set of numbers are known as central tendencies in statistics. In statistics, these acquired numbers are referred to as central or average values. The value of a parameter that is indicative of all the data or the related frequency distribution is called the center or average of any data set or series. Because it illustrates the features or qualities of the complete data set, which would otherwise be exceedingly challenging to notice, such a number is extremely significant. Table 1 shows the definition and mathematical expression of Central Tendency Measure

Table 1: Definition and Mathematical Expression of Central Tendency Measure

S. No	Central Tendency Measure	Definition	Formula
1.	Harmonic Mean	The reciprocal of the average of the reciprocals of the data values	$HM = N/[\left(\frac{1}{B_1}\right), \left(\frac{1}{B_2}\right), \left(\frac{1}{B_3}\right) + .. + \left(\frac{1}{B_N}\right)$ $B_1, B_2, B_3 .. B_N$ are the individual items up to n terms
2.	Median	Center element in the provided data list	-

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

3.	Mode	Range with greater frequency inside a specific dataset	-

3.1.3.2 Degree of Dispersion

By quantifying how far a set of data points diverges from the central tendency, the degree of dispersion sheds light on the distribution or variance of the values. A higher degree of dispersion indicates greater diversity between the data, while greater consistency is suggested by a lower degree. Table 3 shows the definition of Degree of Dispersion and its mathematical expression.

Table 2: Definition of Degree of Dispersion and its mathematical expression.

S. No	Degree of Dispersion	Definition	Formula
1.	Range	The disparity among the highest and lowest values included in a set of data	$Max - Min$
2.	Variance	Average of the squared differences between each data point and the mean of the dataset	$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N}$ N is the number of data points; Y_i is each individual data point; $\bar{Y}$ is the mean of the dataset
3.	SD	Square Root of variance	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N}}$

3.1.4 Feature Selection

From the extracted features the optimal features are selected utilizing the Improved Information Gain. The procedure of selecting a group of relevant features to be utilized in the creation of a framework is termed as feature selection. Figure 8 shows the Feature Selection process.

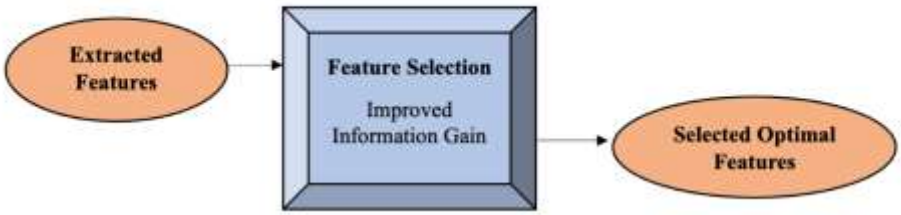

Figure 8: Feature Selection

3.1.4.1 Improved Information Gain

Information gain is one approach that is frequently applied as a term goodness criterion in ML. The information theory states that a phrase gains significance as it adds to the corpus of knowledge. The following defines the knowledge gain of term k:

$$DK(k) = -\sum P(c_i)logP(c_i) + P(k)\sum P\left(\frac{c_i}{k}\right)logP\left(\frac{c_i}{k}\right) + P(\bar{k})\sum P(c_i/\bar{k})logP(c_i/\bar{k}) \quad (9)$$

Where, $P(c_i)$ is the probability of the ith category, $P(D)$ and P $(\bar{D})$ are the probabilities that the term k exists in the documents or not, and J is a collection of documents. In the event that term k appears or does not occur, $P\left(\frac{c_i}{k}\right)$ and $P\left(\frac{c_i}{\bar{k}}\right)$ represent the conditional probability of the ith class value. Unbalances in data classes are resolved by creatively applying balancing factors, which enhances information gathering. This strategy assigns different weights to each class to make that the framework focuses on minority groups and improves general prediction accuracy. Balancing variables helps lessen biases and enables stronger learning, enabling a fairer representation of classes in decision-

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

making. Through the optimization of the recall-precision trade-off, this invention produces a more precise and complete framework. We propose to regulate the conditional entropy percentage of appearing and non-appearing phrases using the balancing factor γ. To find the optimal ω on the unbalanced corpus to improve IG feature selection techniques. Here is how IG with a balance factor are expressed in Eq (4):

$$DK(k,\alpha) = -\sum P(c_i) logP(c_i) + \alpha . P(k) \sum P\left(\frac{c_i}{k}\right) logP\left(\frac{c_i}{k}\right) + (1-\alpha) P(c_i/\bar{k}) logP(c_i/\bar{k}) \quad (10)$$

3.1.5 Deep learning-based Heart Disease Prediction System

The selected optimal features are fed as input to the Deep-learning based heart disease Prediction system which is a combination of CNN, ANN, and RNN. Figure 4 shows the Deep learning-based heart disease detection model.

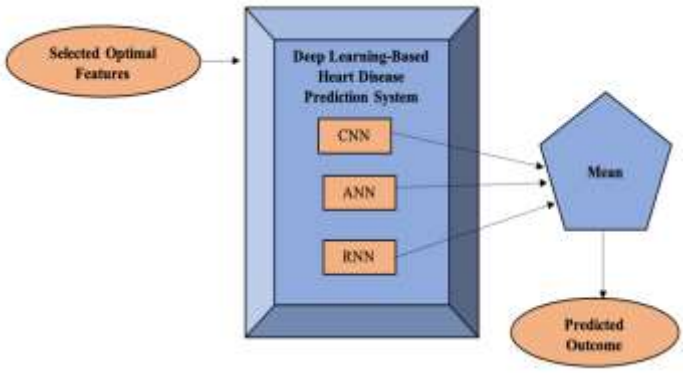

Figure 9: Deep learning-based Heart Disease Prediction System

3.1.5.1 CNN

Using grid-like matrices, convolutional neural networks (CNN), which have evolved from ANN, are mostly used to extract features from datasets. Convolutional, pooling, and fully linked layers are among the layers that make up a CNN in most cases. Convolutional layers perform the work of extracting features from the input data, and the task of down sampling the feature maps by pooling layers reduces their spatial dimensions. The fully linked layers at the network's end carry out classification or regression tasks based on the learned properties. Eq (12) shows the mathematical representation of CNN,

$$x_{i,j}^{l} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} y_{(i+a)(j+b)}^{l-1} \quad (11)$$

3.1.5.2 RNN

DL is a branch of ML and AI that aims to replicate the human brain's analysis of information and learn certain concepts. It is based on neural networks that identify underlying patterns in data collection and are modeled after the brain. Recursive neural networks (RNNs) are subsets of deep neural networks that can learn organized and detailed data by repeatedly using the same set of weights on structured inputs. RNNs are skilled of tackle the graded data due to their indepth tree-like patterns, where parent nodes are produced by connect the child nodes, and each child has the similar weights. To permit for recursive functions and the use of the similar weights, the amount of children for each node in the tree is secure.

$$S = G(\sum_{i=1}^{i=D} V_i D_i) \quad (12)$$

Weight matrices' (V_i) and children's (D_i) products and use the transformation G to determine the parent node's representation.

3.1.5.3 ANN

An ANN is made up of several processing units, usually referred to as neurons, coupled in a specified topology. It can study by doing and generalize from sparse, noisy, and imperfect data. Many data-intensive applications have used ANN with success. A neural network consists of an input layer, several hidden layers, and an output layer. Every layer covers a large number of neurons. Input allocated to it and may obtain any one of 'n' dissimilar inputs. The bias value, 'x_0' is moved to the activation function's input. Let w_1, w_2 , w_3……. w_n be the weights and x_1, x_1, x_2, ……...x_n be the inputs to a neuron. Let 'c' be the bias, and d be the neuron's output, which is calculated as per Eq. (60). The weight of ANN is optimized via Equestrian Ecosystem Optimization (EEO).

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

$$c = A(\sum_{i=0}^{n} w_i x_i + d) \quad (13)$$

Where, A is the activation function employed to retrieve the output from that layer and transfer it as an input to the one below. An ANN is built of nodes and related weights, and these parts often need to be learnt from previously observed patterns. Two types of learning structures are supervised learning and unsupervised learning. The supervised learning network knows what to expect in terms of a response because the output has been labelled.

3.1.5.4 Model Evaluation

Lastly, the hybrid method is assessed using performance metrics such as "accuracy, precision, F-Measure, Sensitivity, Specificity, NPV, MCC, FPR, FNR".

Performance metrics

Several metrics are utilized to calculate the performance counting "Accuracy, Precision, F-Measure, Sensitivity, Specificity, NPV (Negative prediction value), MCC (Matthew's correlation coefficient), FPR (False positive ratio), FNR (False Negative ratio)".

- **Accuracy**

The accuracy is the ratio of properly categorized data to all of the data in the log. The Accuracy is described as,

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (14)$$

- **Precision**

By retaining the whole count of examples used in the classification procedure, precision is the depiction of the sum of genuine examples that are suitably taken into deliberation throughout the prediction method.

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

- **F-Measure**

The definition of the F-measure is the consonant mean of recollect rate and accuracy.

$$F_{Measure} = \frac{2\ Precision \times Recall}{Precision+Recall} \quad (16)$$

- **Sensitivity**

The sensitivity value is gained by just dividing the total positives by the proportion of true positive predictions.

$$Sensitivity = \frac{TP}{TP+FN} \quad (17)$$

- **Specificity**

Specificity is premeditated by dividing the count of accurately anticipated negative outcomes by the sum of negatives.

$$Specificit\ y\ = \frac{TN}{TN+FP} \quad (18)$$

- **NPV**

NPV determines the efficiency of an analytical test or another quantifiable metric.

$$NPV = \frac{TN}{TN+FN} \quad (19)$$

- **MCC**

Below is a depiction of the two-by-two binary variable connotation measure recognized as MCC,

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+FN)(TN+FP)(TN+FN)(TP+FP)}} \quad (20)$$

- **FPR**

The count of negative occurrences divided by the count of negative events that were erroneously ranked as positive yields the false positive rate (false positives).

$$FPR = \frac{FP}{FP+TN} \quad (21)$$

- **FNR**

The probability that an actual positive may be ignored by the trial is termed as the false-negative rate, occasionally mentioned to as the "miss rate".

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

$$FNR = \frac{FN}{FN+TP} \quad (22)$$

4. Result and Discussion

This research suggests a DL-based HDPS which is a combination of ANN, CNN and RNN. Also, a Novel Improved Information Gain used to select the applicable features. In this segment, the suggested approach is compared with different current technique to know the performance efficiency of recommended method.

4.1 Overall Performance Analysis: Comparative Analysis and Discussion (For Training rate=70)

The table 3 presents the evaluation metrics for different models, including Precision, Recall, F-Measure, Matthews Correlation Coefficient (MCC), Sensitivity (Sen), Specificity (Spec), Accuracy, Negative Predictive Value (NPV), False Positive Rate (FPR), and False Negative Rate (FNR). These metrics are crucial for assessing the performance of the Heart Disease Prediction System (HDPS) and comparing it with other models such as Convolutional Neural Networks (CNN), Recursive Neural Networks (RNN), Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Multilayer Perceptron (MLP). The proposed HDPS demonstrates outstanding performance across all evaluation metrics. It achieves high Precision (0.988), Recall (0.996), and F-Measure (0.9979), indicating its ability to accurately identify instances of heart disease while minimizing false positives and false negatives. The MCC score of 0.9881 further confirms the robustness of the HDPS in handling imbalanced datasets and providing reliable predictions. Comparatively, the CNN, RNN, ANN, LSTM, and MLP models also exhibit commendable performance, albeit slightly lower than the proposed HDPS. The HDPS incorporates a hybrid architecture that combines Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recursive Neural Networks (RNN). This amalgamation allows the model to capture a broader range of features from diverse data sources such as structured, imaging, and sequential data. By leveraging multiple neural network architectures, the HDPS can effectively extract intricate patterns and relationships inherent in heart disease data, leading to more accurate predictions compared to single-model approaches. The HDPS is designed to handle the complexities and nuances of heart disease prediction through rigorous preprocessing, feature extraction, and feature selection techniques. By cleaning the data, handling missing values, and selecting informative features using the Improved Information Gain technique, the HDPS mitigates the impact of noisy and irrelevant data, thereby enhancing model robustness and generalization capabilities. This robustness ensures consistent and reliable performance across diverse datasets and scenarios, making the HDPS more suitable for real-world applications.

Table 3: Result overall values of metrices

	Sen	Spec	Acc	Precision	Recall	F-Measure	NPV	FPR	FNR	MCC
Proposed	0.988	0.996	0.9979	0.9881	0.9881	0.9881	0.996	0.004	0.0119	0.9841
CNN	0.976	0.9922	0.9884	0.9767	0.9767	0.9767	0.992	0.0078	0.0233	0.9689
RNN	0.972	0.9907	0.9861	0.9722	0.9722	0.9722	0.99	0.0093	0.0278	0.963
ANN	0.97	0.9903	0.9854	0.9708	0.9708	0.9708	0.99	0.0097	0.0292	0.961
LSTM	0.959	0.9865	0.9797	0.9594	0.9594	0.9594	0.986	0.0135	0.0406	0.9458
MLP	0.951	0.9838	0.9757	0.9514	0.9514	0.9514	0.983	0.0162	0.0486	0.9352

According to the data provided in Table 3, a recommended method achieved the highest rating across all assessed criteria. As a result, the proposed model outperforms all other methods that were evaluated.

4.2 Classifier Performance Analysis

Additionally, a line graph representing each metrics evaluation (in terms of classifier performance for heart disease detection) is shown in Figures 10 through 19. Specificity is an especially crucial parameter in calculation. It has been contrasted with multiple current and suggested algorithms. The Specificity score is shown in Figure 10. The graph in Figure 10 above was made using the numbers from Table 3. In this study, the specificity values for CNN, RNN, ANN, LSTM and MLP the Proposed model are presented as (0, 9922), (0.9907), (0.9903), (0.9865) (0.9838) and (0.996) respectively. The suggested approach is the most sensitivity available. The newly developed model offers a high degree of sensitivity in contrast to earlier methods that are presently in utilization.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

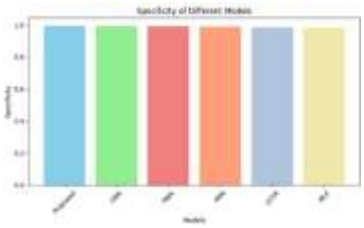

Figure 10: Analysis of Specificity

The results of comparing a proposed approach's Accuracy with other techniques are shown graphically in Figure 11 below. The values from Table 3 were used to create table above graph in Figure 11. The values of Accuracy for CNN, RNN, ANN, LSTM and MLP, and Proposed model in this investigation are given as (0.9884), (0.9861), (0.9854), (0.9797) and (0.9979) correspondingly. The proposed method is the most precise of all the options. In comparison to previous models that are currently in use, the newly created model has a high level of Accuracy.

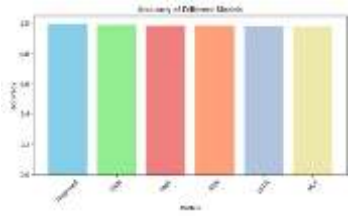

Figure 11: Analysis of Accuracy

The values from Table 3 were used to create table above graph in Figure 12. The values of Precision for CNN, RNN, ANN, LSTM and MLP, and Proposed model in this investigation are given as (0.9767), (0.9722), (0.9708), (0.9594), (0.9514) and (0.0.9881) correspondingly. The proposed method is the most precise of all the options. In comparison to previous models that are currently in use, the newly created model has a high level of Accuracy.

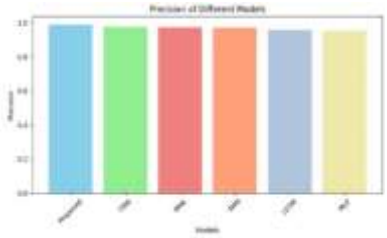

Figure 12: Analysis of Precision

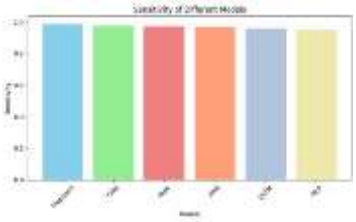

Figure 13: Analysis of Sensitivity

The values from Table 3 were used to create table above graph in Figure 13. The values of Sensitivity for CNN, RNN, ANN, LSTM and MLP, and Proposed model in this investigation are given as (0.976), (0.972), (0.97), (0.959), (0.951), and (0.988) correspondingly. The proposed method is the most precise of all the options. In comparison to previous models that are currently in use, the newly created model has a high level of Sensitivity. The Recall of a suggested approach have been compared with other existing methods, and outcomes are visually displayed in Figure 14 below.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

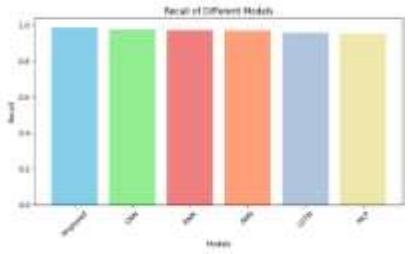

Figure 14: Analysis of Recall

The graph in Figure 14 above was created using the data from Table 3. CNN, RNN, ANN, LSTM, MLP and the suggested model all have Recall values of (0.9767), (0.9722), (0.9708), (0.9594), (0.9514) and (0.9881) respectively. The best Recall score is obtained using a suggested approach. The designed model used in this study offers the greatest Recall compared to the previous ones that are still in use.

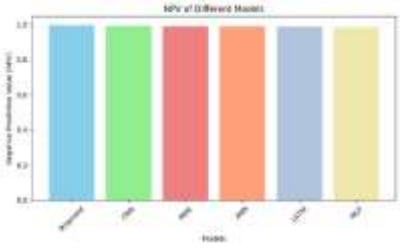

Figure 15: Analysis of NPV

The graph in Figure 15 above was created using the data from Table 3. CNN, RNN, ANN, LSTM, MLP and the suggested model all have Recall values of (0.992), (0.99), (0.99), (0.986), (0.983) and (0.996) respectively. The best Recall score is obtained using a suggested approach. The designed model used in this study offers the greatest NPV compared to the previous ones that are still in use. Figure 16 below illustrates the results of comparing the MCC of a proposed model to those of competing strategies.

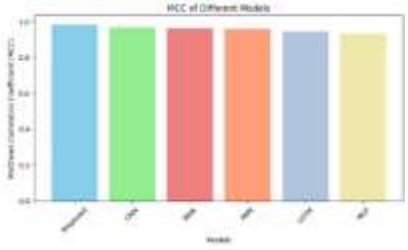

Figure 16: Analysis of MCC

The information from Table 3 was used to construct the graph in Figure 16 above. The MCC values for CNN, RNN, ANN, LSTM, MLP and the proposed model are (0.9689), (0.963), (0.961), (0.9458), (0.9352) and (0.9841) respectively. Since the suggested technique has the highest value, it is a most efficient method than other models.

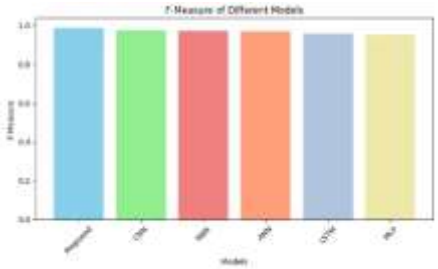

Figure 17: Analysis of F-Measure

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

The data from Table 3 was used to construct the graph in Figure 17 above. The F-Measure values for CNN, RNN, ANN, LSTM, MLP and the proposed model are (0.9767), (0.9722), (0.9708), (0.9594), (0.9514) and (0.9881) respectively. Since the suggested technique has the highest value, it is a most efficient method than other models.

The outcomes of contrasting the FPR of a proposed plan to those of competing strategies are shown in Figure 18 below.

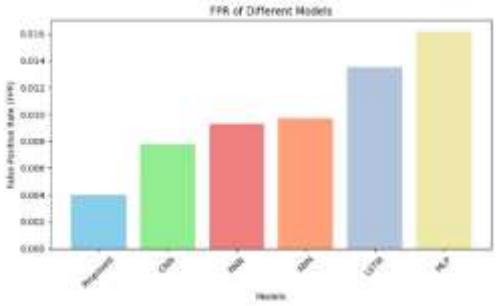

Figure 18: Analysis of FPR

The data from Table 3 was utilized to construct the graph in Figure 18 above. The FPR values for CNN, RNN, ANN, LSTM, MLP and the proposed model are (0.0078), (0.0093), (0.0097), (0.0135), (0.0162) and (0.004) respectively. The proposed method yields a lowest FPR and FNR value. In comparison to the earlier models that are still in use, the created model employed in this study has the lowest FPR rate. The model with least FPR value will perform more efficiently. Since the suggested technique has least value, it is a most efficient method than other models.

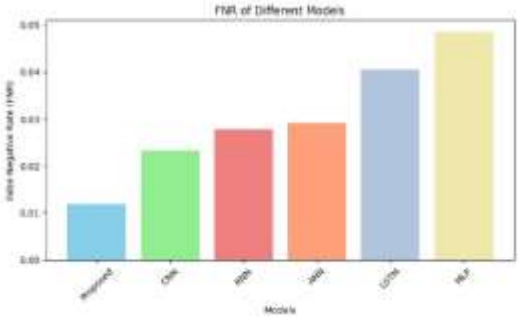

Figure 19: Analysis of FNR

The graph shown in Figure 19 above was created using the information from Table 3. CNN, RNN, ANN, LSTM, MLP, and the suggested model have the following FNR values: 0.0233, 0.0278, 0.0292, 0.0406, 0.0486, and 0.0119, correspondingly. The suggested approach produces the lowest FNR value. The developed model used in this study has the lowest FNR rate when compared to the older models that are still in use. More efficiency will be shown by the model with the lowest FNR value. Compared to other models, the recommended strategy is the most efficient because it has the lowest value.

4.3Analysis on Encryption and Decryption Time

Table 4 provides a comparative analysis of encryption and decryption times for various encryption techniques, including the proposed model with a two-level security mechanism combining Blowfish and RSA algorithms, as well as other commonly used encryption algorithms such as ECC, AES, and ElGamal. The proposed model exhibits the shortest encryption time of 3 minutes and 15.24 seconds, outperforming all other encryption techniques. This efficiency in encryption time can be attributed to the optimized combination of Blowfish and RSA algorithms, which efficiently encrypt data while ensuring robust security measures. Similarly, the proposed model also demonstrates the fastest decryption time of 1 minute and 32.88 seconds. This rapid decryption process is crucial for real-time applications where quick access to encrypted data is essential for decision-making and analysis. In comparison, other encryption techniques such as ECC, AES, and ElGamal exhibit longer encryption and decryption times. ECC requires 5 minutes and 22.08 seconds for encryption and 2 minutes and 38.5 seconds for decryption. AES and ElGamal also show relatively longer encryption and decryption times, further emphasizing the superior efficiency of the proposed

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

model. Overall, the encryption and decryption times of the proposed model highlight its effectiveness in balancing security and performance requirements. By leveraging a two-level security mechanism combining Blowfish and RSA algorithms, the proposed model achieves rapid encryption and decryption times without compromising data security, making it well-suited for secure and efficient data transmission and storage applications. The proposed model's flexibility allows it to handle diverse data types and sizes efficiently. Whether encrypting small text files or large multimedia streams, the model can adapt to the specific requirements of different data formats. This adaptability ensures consistent performance across various applications and use cases, making it a versatile solution for secure data storage and transmission.

Table 4: Encryption and Decryption Time of the proposed model.

	Encryption Time(min/sec)	**Decryption Time(min/sec)**
Proposed	3.254	1.548
ECC	5.368	2.635
AES	5.965	2.942
ELGAMAL	6.358	3.687

The proposed model outperforms Elliptic Curve Cryptography (ECC), Advanced Encryption Standard (AES), and ELGAMAL in terms of encryption and decryption times, with a significantly shorter time of 3.254 minutes for encryption and 1.548 minutes for decryption. This makes it a more efficient choice for data security, making it a more time-effective choice compared to established encryption techniques. The shorter times indicate the model's efficiency in securing data. Figure 20 shows the graphical representation of the Encryption and Decryption Time.

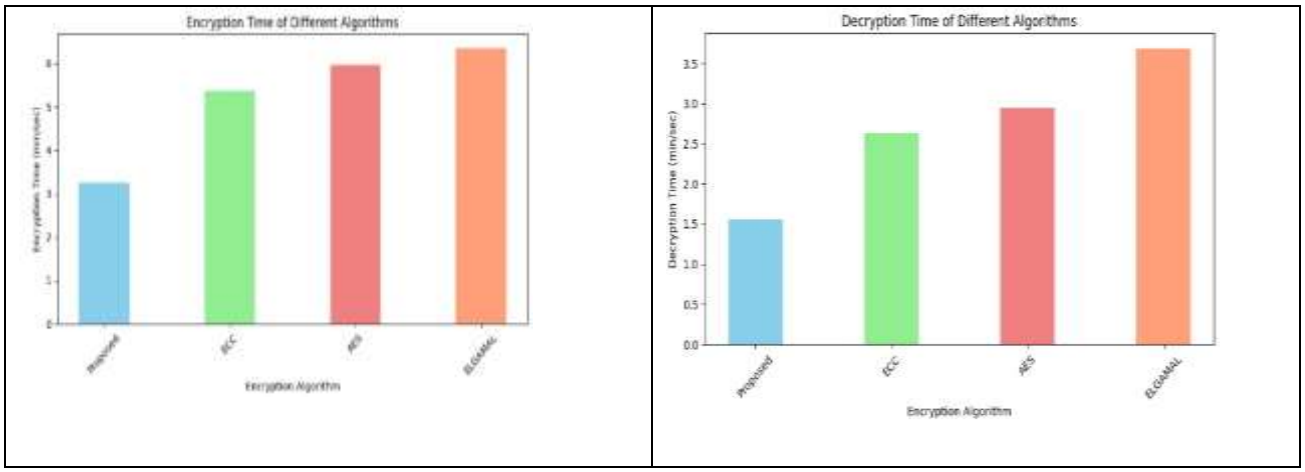

Figure 20: Graphical Representation of the Encryption and Decryption Time.

5. Conclusion

The proposed privacy-preserving scheme addresses critical concerns surrounding data privacy and security in cloud-based heart disease prediction. By integrating encryption techniques with deep learning models, the system ensures secure data transmission and storage while maintaining high predictive accuracy. The results highlight the efficacy of the proposed approach in achieving superior performance metrics. Future research could focus on scalability, interoperability with electronic health record systems, and incorporation of advanced encryption protocols for enhanced data protection. Overall, the proposed scheme offers a promising solution for privacy-preserving heart disease prediction in modern healthcare systems.

Key Outcomes:

1. **Performance**: The proposed encryption scheme demonstrates superior performance in terms of encryption and decryption times compared to other encryption techniques such as ECC, AES, and ElGamal. With encryption

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

times of 3.254 minutes and decryption times of 1.548 minutes, the proposed scheme offers efficient data protection without compromising speed.

2. **Accuracy**: In heart disease prediction, the proposed model achieves exceptional performance with a precision of 0.988, recall of 0.996, and an F-measure of 0.9979. These metrics indicate the model's ability to accurately identify instances of heart disease while minimizing false positives and false negatives.

3. **Security**: By incorporating a two-level security mechanism combining RSA and Blowfish encryption techniques, the proposed scheme ensures robust data security in cloud-based healthcare systems. This enhances patient privacy and safeguards sensitive medical information against unauthorized access.

Future Scope:

- Enhanced Security: Continued research can explore homomorphic encryption for computations directly on encrypted data, further strengthening privacy.
- Federated Learning: Investigating federated learning approaches would enable distributed training without data sharing, addressing privacy concerns.
- Scalability and Real-Time: Addressing scalability and real-time processing challenges will become crucial for large-scale deployments.

ACKNOWLEDGEMENT: The author extends the appreciation to the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for funding this research work through the project number (R-2024-10xy).

Reference

[1] Yang, H., Luo, Y.M., Ma, C.Y., Zhang, T.Y., Zhou, T., Ren, X.L., He, X.L., Deng, K.J., Yan, D., Tang, H. and Lin, H., 2023. A gender specific risk assessment of coronary heart disease based on physical examination data. *NPJ digital medicine*, *6*(1), p.136.

[2] Ali, M.M., Paul, B.K., Ahmed, K., Bui, F.M., Quinn, J.M. and Moni, M.A., 2021. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, *136*, p.104672.

[3] Shah, D., Patel, S. and Bharti, S.K., 2020. Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*, pp.1-6.

[4] Wagle, A.A., Isakadze, N., Nasir, K. and Martin, S.S., 2021. Strengthening the learning health system in cardiovascular disease prevention: time to leverage big data and digital solutions. *Current atherosclerosis reports*, *23*, pp.1-8.

[5] Tiwari, P., Colborn, K.L., Smith, D.E., Xing, F., Ghosh, D. and Rosenberg, M.A., 2020. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. *JAMA network open*, *3*(1), pp.e1919396-e1919396.

[6] Desai, F., Chowdhury, D., Kaur, R., Peeters, M., Arya, R.C., Wander, G.S., Gill, S.S. and Buyya, R., 2022. HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing. *Internet of Things*, *17*, p.100485.

[7] Faridi, F., Sarwar, H., Ahtisham, M. and Jamal, K., 2022. Cloud computing approaches in health care. *Materials Today: Proceedings*, *51*, pp.1217-1223.

[8] Abdelfattah, S., Badr, M.M., Mahmoud, M., Abualsaud, K., Yaacoub, E. and Guizani, M., 2023. Efficient and privacy-preserving cloud-based medical diagnosis using an ensemble classifier with inherent access control and micro-payment. *IEEE Internet of Things Journal*.

[9] Ren, W., Tong, X., Du, J., Wang, N., Li, S.C., Min, G., Zhao, Z. and Bashir, A.K., 2021. Privacy-preserving using homomorphic encryption in Mobile IoT systems. *Computer Communications*, *165*, pp.105-111.

[10] Ali, A., Al-Rimy, B.A.S., Alsubaei, F.S., Almazroi, A.A. and Almazroi, A.A., 2023. HealthLock: Blockchain-Based Privacy Preservation Using Homomorphic Encryption in Internet of Things Healthcare Applications. *Sensors*, *23*(15), p.6762.

[11] Li, D., Liao, X., Xiang, T., Wu, J. and Le, J., 2020. Privacy-preserving self-serviced medical diagnosis scheme based on secure multi-party computation. *Computers & Security*, *90*, p.101701.

[12] Khan, J.A., 2024. Role-Based access Control (RBAC) and Attribute-Based Access Control (ABAC). In *Improving Security, Privacy, and Trust in Cloud Computing* (pp. 113-126). IGI Global.

[13] Kanwal, T., Anjum, A. and Khan, A., 2021. Privacy preservation in e-health cloud: taxonomy, privacy requirements, feasibility analysis, and opportunities. *Cluster Computing*, *24*, pp.293-317.

[14] Padinjappurathu Gopalan, S., Chowdhary, C.L., Iwendi, C., Farid, M.A. and Ramasamy, L.K., 2022. An efficient and privacy-preserving scheme for disease prediction in modern healthcare systems. *Sensors*, *22*(15), p.5574.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024

[15] Jayaram, R. and Prabakaran, S., 2021. Onboard disease prediction and rehabilitation monitoring on secure edge-cloud integrated privacy preserving healthcare system. *Egyptian Informatics Journal*, *22*(4), pp.401-410

[16] Munirathinam, T., Ganapathy, S. and Kannan, A., 2020. Cloud and IoT based privacy preserved e-Healthcare system using secured storage algorithm and deep learning. *Journal of Intelligent & Fuzzy Systems*, *39*(3), pp.3011-3023.

[17] Xu, C., Wang, N., Zhu, L., Zhang, C., Sharif, K. and Wu, H., 2021. Reliable and privacy-preserving top-k disease matching schemes for E-healthcare systems. *IEEE Internet of Things Journal*, *9*(7), pp.5537-5547.

[18] Khan, M.A., Abbas, S., Atta, A., Ditta, A., Alquhayz, H., Khan, M.F. and Naqvi, R.A., 2020. Intelligent Cloud Based Heart Disease Prediction System Empowered with Supervised Machine Learning. *Computers, Materials & Continua*, *65*(1).

[19] Nancy, A.A., Ravindran, D., Raj Vincent, P.D., Srinivasan, K. and Gutierrez Reina, D., 2022. Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics*, *11*(15), p.2292.

[20] Li, D., Liao, X., Xiang, T., Wu, J. and Le, J., 2020. Privacy-preserving self-serviced medical diagnosis scheme based on secure multi-party computation. *Computers & Security*, *90*, p.101701.

[21] Wang, W., Li, X., Qiu, X., Zhang, X., Brusic, V. and Zhao, J., 2023. A privacy preserving framework for federated learning in smart healthcare systems. *Information Processing & Management*, *60*(1), p.103167.

[22] Sharma, S., Chen, K. and Sheth, A., 2018. Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Computing*, *22*(2), pp.42-51.

[23] Chenthara, S., Ahmed, K., Wang, H. and Whittaker, F., 2019. Security and privacy-preserving challenges of e-health solutions in cloud computing. *IEEE access*, *7*, pp.74361-74382.

Doi: https://doi.org/10.54216/FPA.150109
Received: August 19, 2023 Revised: December 16, 2023 Accepted: February 19, 2024