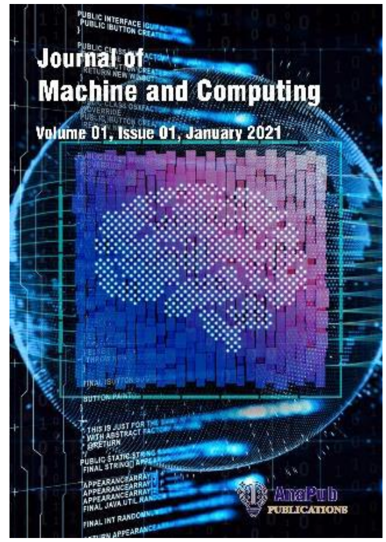
Journal Pre-proof



Integrating Homomorphic Encryption with Blockchain Technology for Machine Learning Applications

Subhra Prosun Paul, Sreenivasu S V N, Md. Shafikul Islam, Raghunath B, Kanchan Dhote and Vetrithangam D

DOI: 10.53759/7669/jmc202505031

Reference: JMC202505031

Journal: Journal of Machine and Computing.

Received 24 March 2024

Revised form 16 June 2024

Accepted 27 November 2024

Please cite this article as: Subhra Prosun Paul, Sreenivasu S V N, Md. Shafikul Islam, Raghunath B, Kanchan Dhote and Vetrithangam D, "Integrating Homomorphic Encryption with Blockchain Technology for Machine Learning Applications", Journal of Machine and Computing. (2025). Doi: https://doi.org/10.53759/7669/jmc202505031

Integrating Homomorphic Encryption with Blockchain Technology for Machine Learning Applications

[1]Subhra Prosun Paul, [2]S. V. N Sreenivasu, [3]Md. Shafikul Islam, [4]B. Raghunath, [5]Kanchan Dhote, [6]D. Vetrithangam

[1]Department of CSE, Uttara University, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Narasaraopeta Engineering College, Andhra Pradesh, India
[3]Department of CSE, Uttara University, Dhaka, Bangladesh
[4]Department of EEE, Sri Manakula Vinayagar Engineering College, Madagadipet, Pondicherry, India
[5]Department of Electronics and Computer Science, Shri Ramdeobaba College of Engineering and Management, Ramdeobaba University, Nagpur, Maharashtra, India.
[6]Department of CSE, Chandigarh University, Mohali, Punjab, India

subhra.phd.cu2021@gmail.com, drsvnsrinivasu@gmail.com, shafikul.islam@uttarauniversity.edu.bd, raghushara1@gmail.com, hdhotek@rknec.edu, vetrigold@gmail.com

Abstract - Leveraging cutting-edge technology like blockchain and machine intelligence, smart healthcare systems have emerged as a potential strategy for enhancing healthcare services. In order to secure health data, this study offers a unique design and analysis of a smart healthcare system that applies blockchain technique and the Paillier homomorphic encryption algorithm in addition to a machine learning algorithm to detect cardiological disease. The suggested method seeks to solve the problems with predictive analytics and safe health data exchange in the medical field. Sensitive information is encrypted during transmission and storage using the Paillier Homomorphic Encryption technique, guaranteeing its confidentiality. By providing traceability and accountability in data access and sharing, blockchain technology is used to construct a safe and transparent record of health transactions. In addition, a machine learning algorithm is used to forecast cardiac illness based on the encrypted data, giving medical practitioners insightful information to help them make judgments. The integration of these technologies and their advantages in improving healthcare services are highlighted in the discussion of the proposed scheme's constructional and operational specification section. Simulation experiments are used to assess the suggested method's efficiency and reflect its efficacy in terms of data security, detection accurateness, and computing proficiency. Comparing the integrated approach to conventional approaches, the results demonstrate a considerable improvement in prediction accuracy and security of health data. To sum up, the suggested smart healthcare system provides a thorough approach to guaranteeing the security of patient data and enhancing predictive analytics in the medical field. Machine learning, blockchain technology, and Paillier homomorphic encryption are all integrated into it, which shows promise for improving healthcare services and developing the field of smart healthcare systems.

Keywords – ***Blockchain Technology, Machine Learning, Homomorphic Encryption, Accuracy, Healthcare System, Internet of Medical Things (IoMT)***

I. INTRODUCTION

Technology and its use is growing at a very quick pace, particularly in the medical field. We can create a smart healthcare system using numerous techniques like machine learning (ML), cryptography and so on [1]. Using IoT, the internet, sensors, actuators, and other devices, the smart healthcare system (SHS) records, analyses, and shares patient data to continuously monitor patients and provide useful health insights to medical professionals for better treatment [2]. SHS aims to provide more efficient, convenient, and customized care by updating current healthcare systems. Smart thermometers and wearable biosensors track health data, such as blood sugar levels, to give patients and healthcare professionals individualized insights. Smart thermometers that continuously monitor body temperature can help to detect infections early and ensure that patients receive timely medical attention. Other benefits of tele-health services and virtual hospitals include the ability for medical professionals to provide remote diagnosis, consultation, and treatment via digital tools and video conferencing. SHS elements include electronic health records, IOT-enabled medical equipment, patient involvement, data analytics, predictive modeling, telemedicine, health information sharing, and healthcare automation [3].

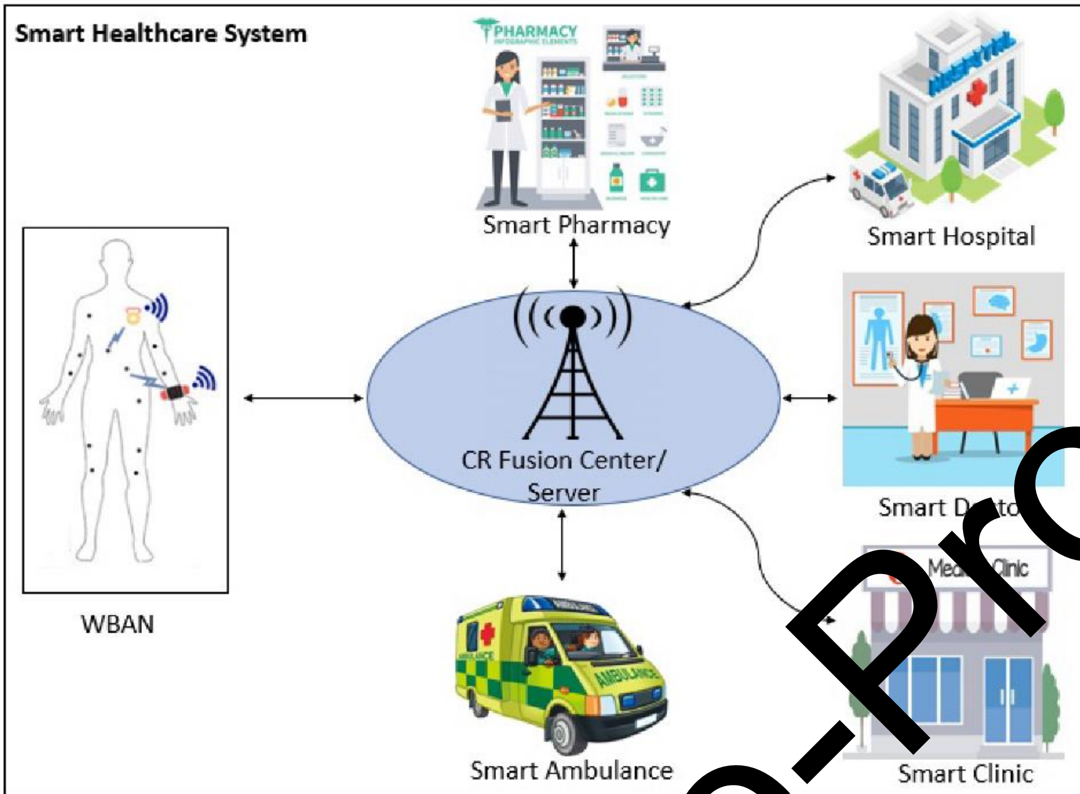


Fig 1. Smart Healthcare System

Figure 1 show the physical infrastructure of smart healthcare system (SHS) where smart hospital, smart doctor, smart clinic, smart ambulance, and smart pharmacy are connected with server. The medical server collects health data from human body which uses wireless body area network (WBAN). Different wearable and implantable medical sensor devices are connected with this WBAN.

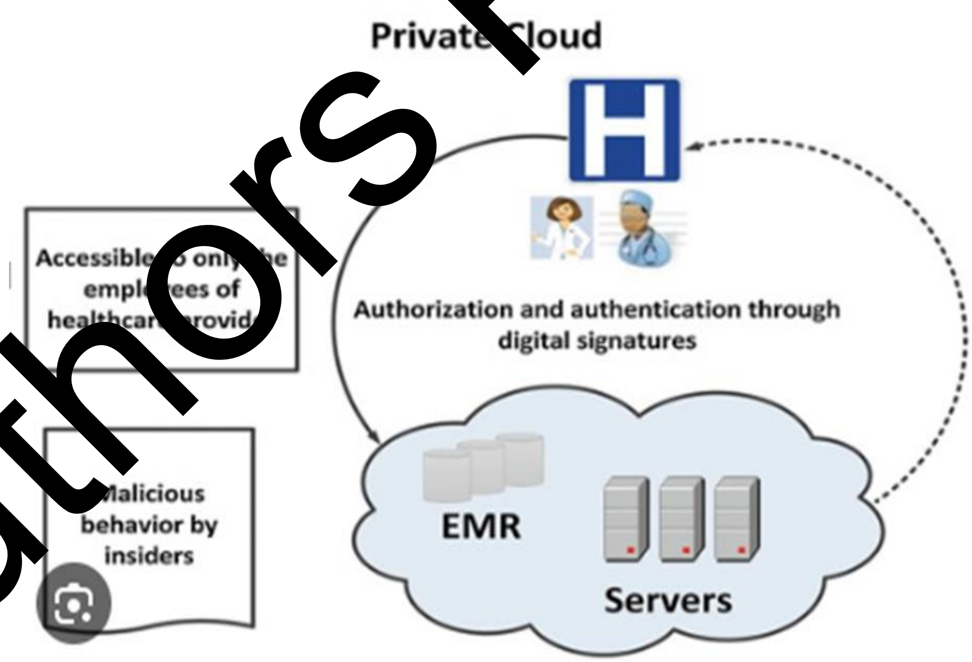


Fig 2. Cloud-Based Healthcare System

Figure 2 shows cloud based healthcare system. Smart healthcare system uses various smart devices and technologies to connect patients, doctors, hospitals, and pharmacies.AI algorithms for diagnosis and decision support, cloud computing infrastructure; medical imaging technologies like MRIs and CT scanners, health sensors and IoT devices, etc. are the indispensable elements of

SHS [4][34]. These devices use IoMT to assist in gathering and sending patient health data in smart healthcare applications. Intelligent medical technology (IMT) devices, which include wearable, implantable medical sensors, serve as smart assistants in the healthcare industry [5]. They facilitate communication between patients and physicians while ensuring their well-being. Pharmaceutical management, chronic illness management, clinical workflow optimization, hospital asset management, and health wellbeing tracking are all use of IoMT devices [6].

A branch of artificial intelligence (AI) called ML aids in data analysis and summarization of findings. In short, ML is the capacity of a machine that can imitate human cognitive ability. Mathematical model mapping techniques called ML algorithms are employed to find underlying outlines hidden in data. For various applications, such as regression, classification, prediction, clustering, etc., there exist various machine learning methods [7]. Medical algorithms fall into three main categories predictive, diagnostic, and machine learning. While predictive and diagnostic algorithms help to confirm problems, machine learning algorithms use data to find patterns for early identification and handling approaches. The application of ML in healthcare includes the following: medication development, hospital management optimization, health insurance, virtual nursing, medical imaging, disease outbreak prediction, patient behavior modification, accurate diagnostics, and identification of high-risk patients [8].

Fig 3. Usage of Machine Learning (ML) in Healthcare System

Figure 3 represents various usage of ML technique in healthcare system which is referred from https://www.rishabhsoft.com.Machine learning helps by making diagnoses by analyzing images like X-rays and MRIs, helping with research by using large amounts of medical data, and supporting doctors with tasks like identifying patients at risk of certain diseases. ML models can be trained with historical data to guess a patient's risk of contracting a concrete disease based on a mix of variables like life leading process or genetics [9]. Naïve Bayes classifier, a supervised ML technique, is used to forecast the disease. The Naïve Bayes method determines the likelihood of the illness. Classifier for decision trees used to assess the model. End users make use of this system [10]. The algorithm will use symptoms to forecast illness. The technology used by this system is machine learning. An ensemble of decision trees, each trained on data subsets with random feature selection, is used by the Random Forest algorithm. Each decision tree in the "V mechanism" makes predictions, and the result is decided by majority vote. The ultimate forecast is made by the class that received the most votes. Furthermore, the system uses feature significance to gauge how useful a feature is for producing precise predictions. Random feature selection and bagging are used to reduce over fitting and enhance generalization. After being trained on labeled data, the algorithm uses the patterns it has learned to predict the results of fresh cases [11].

Now a day, Internet of Medical Things (IoMT) an exceptional type of IoT network is implemented in healthcare field where privacy and security are the key concern to be handled. With the help of block chain and machine learning technique, healthcare fifth generation system is developed into smart healthcare system. The basic objective of smart healthcare is to reduce patient stress and healthcare cost. Machine learning is such a technique where a central server is used in IoMT network [12]. In this paper, a smart healthcare system is designed using block chain technology and Paillier homomorphic encryption for privacy and security of health data where an intrusion detection system will detect any unauthorized activities in the healthcare network. Random Forest algorithm is also being used to predict disease (heart) in this SHS. Doctors can monitor patient's condition continuously using various medical sensors which will be attached with patients and will be connected to the IoMT network. These sensor will collect patients medical information like temperature, heart bit, BP etc. from time to time and can transfer these information to doctor so that doctor can take necessary medical step for the patients through this smart healthcare system. In this proposed framework, several numbers of medical organizations are connected through local model which is directly connected to a global model so that all the clinical data can be shared very secretly throughout the IoMT based smart healthcare network.

Confidentiality and Safety Analysis of Health Data in SHS

It is crucial to protect the safety and confidentiality of medical data in smart healthcare systems. A number of important tactics may be used to accomplish this. Data encryption is essential first and foremost. Sensitive information can be shielded from interception even in the event that it is intercepted by employing robust encryption methods like AES. Control techniques for access are also crucial. Confirming that only authorized individuals have access to the confidential data is facilitated by the application of stringent access controls, such as multi-factor authentication (MFA) and role-based access controls (RBAC). Data reduction techniques should also be used, gathering and retaining just the information required running the smart healthcare system. By doing this, the chance of a data leak is decreased [13]. The identity of the people whose data is anonymized or pseudonymized can be further protected, providing an additional degree of privacy protection. Data in transit should be encrypted during transmission and protected by using secure communication protocols like HTTPS, TLS, and VPNs. In order to identify and hit system susceptibilities, regular safety audits and penetration tests should be done. To guarantee the integrity of the data and identify any unauthorized modifications, data integrity tests, such as digital signatures or checksums, should be used [14].

To safeguard the safety and confidentiality of health data, it is also essential to ensure compliance with all applicable laws, regulations, and standards, including GDPR, HIPAA, and others. Comprehensive privacy and security strategy for smart healthcare systems also must include training users and staff on best practices for data privacy and security and having an incident reaction strategy in place to react quickly and to alleviate any safety incidents or data cracks [15]. Given the confidentialness nature of this data, secrecy and safety analysis of medical or health data in smart healthcare systems is essential. Here are some crucial ideas and methods to keep in mind:

1. Data Encryption
2. Access Control
3. Data Minimization
4. Protected Communication Protocol
5. Safety Audit and Penetration Testing
6. Data Integrity Checking

These procedures may be added to SHS to improve the safety and confidentiality of medical and health data, guaranteeing that patient data is shielded from breaches and unwanted access [16].

Research Objectives:

1. To diagnose patients medical data and to provide effective treatment on real time basis remotely.
2. To transfer real time medical data accurately to the medical organization so that highest level of effective treatment can be ensured.
3. To provide a secured and intrusion detection based data transmission mechanism for SHS network based smart healthcare system.

II. BACKGROUND STUDY

A great portion of data is often generated in the medical field. But it's frequently not used correctly. The data suggests that there are some underlying patterns and their relationships in the created text, image, sound, or file. Manually analyzing medical data is a difficult and time-consuming process. ML enters the picture here, enabling our task (i.e., analyzing the medical data) with ease. Then vast medical information can be analyzed and predictions may be made with the help of various types of ML methods [17]. Some of such algorithms and their applications to medical diagnostics in the medical field will be shielded in this section.

One of the key applications of ML algorithms in the medical area is the prediction of a patient's risk of heart disease, particularly based on characteristics like gender, blood pressure, cholesterol, and stress [18]. It is critical to precisely and punctually detect heart disease to treat people when it is needed. The risk of heart disease is determined using data mining methods like logistic regression, AdaBoost, Naive Bayes, decision trees, and support vector machines. One of the most fatal cardiac conditions is cardiopathy, which can be inexpensively detected using ML algorithms. Publicly available datasets like CHSLB (Cleveland, Hungary, Switzerland, and Long Beach) are used to assess the efficacy of the ML prototypes.in the prognosis of cardiac disease [19].

IoT makes it easier for individuals to stay connected to products and people in their daily lives. By integrating cloud technologies, we can make IoT devices function more efficiently. Cloud-integrated Internet of Things devices facilitate data collection, edge computing implementation, early illness prediction, prompt reactions from medical professionals, and effective service delivery [20]. By eliminating pointless data from the dataset, feature selection facilitates a faster model training procedure and expedites the model's convergence stage. Fast Conditional Mutual Information (FCMIM) is a feature selection technique that chooses features by utilizing conditional mutual information. When combined with ML classifiers such as SVM, the FCMIM algorithm allows us to obtain predictions that are more accurate than those made with neural networks. To assess how well machine learning models are performing, we can apply cross-validation (C-V) approaches. In C-V, the dataset is broken into training and test sets more than once. Rather, the dataset is regularly divided into smaller groups, and the performance in each group is then averaged. In doing so, we lessen the effect of partition randomness on the outcome [21]. Predicting diabetes can also benefit from machine learning. Diabetes is linked to numerous other illnesses. If detected early on, the disease's negative consequences can be avoided. Additionally, a lot of data analysts are working to create an ML model that can predict diabetes more precisely [22].

There are some models that improve model performance and lower error rates, such as the Intelligent Diabetes Mellitus Prediction Framework (IDMPF). One of the structured datasets that is made available to the public and is used to forecast diabetes is called PIDD (Pima India Diabetes Database). The primary variables that we take into account when predicting diabetes are age, BMI, insulin, pregnancies, skin thickness, and blood glucose levels. The accuracy of logistic regression in predicting diabetes is 86%. According to certain study, combining ML algorithms with IOT-edge-cloud computing yields superior outcomes. For example, when comparing the Random forest ML algorithm's performance with logistic regression, it does well in predicting diabetes among the PIMA Indian and Sylhet datasets [23]. The human liver is related with a vast amount of data, therefore diagnosing liver illnesses is a difficult undertaking. ML aids in the prognosis of liver disease. The Indian Liver Patient Dataset, which was acquired by UCI ML Repository, is one of the frequently used datasets for this endeavor. We can get better outcomes by combining ML classifier algorithms like SVM, RF, and Decision tree with AI techniques like genetic algorithms and particle swarm optimization [24].

Promising developments in predictive diagnoses for chronic renal disease are possible with the use of ML in the medical area. Through the application of preprocessing techniques such as data normalization, missing value handling, and category to numerical conversion, we may enhance model functionality and lower the error rate [25]. In order to refine the datasets and identify important predictors such as age, haemoglobin etc. There isn't a perfect machine learning model for predicting diseases. To acquire better results, we need to experiment with machine learning algorithms and apply various cross-validations and feature selection strategies depending on the dataset [26].

Findings of the Existing Research

1. There is no efficient data mining methods to justify the clinical data which is used to detect disease and to prescribe properly. Because, this data is very sensitive.
2. Patient’s privacy is not maintained properly. Because patients medical data is shared throughout the SHS network where several medical organizations are connected both locally and globally.
3. There is no threat handling mechanism during the clinical data transmission throughout the SHS network.

III. PROPOSED ARCHITECTURE

The following figure 4 shows the proposed architecture of paillier homomorphic encryption and blockchain (IPFS) technology based secured smart healthcare system. In this architecture, paillier homomorphic encryption technique is implemented while data is transferred from session layer to gateway layer. From the gateway layer, the encrypted data is sent to the storage layer where hash value is stored in the form of IPFS. After that these encrypted data is trained through ML algorithm (Random Forest) to get the predicted result. These encrypted (predicted) results are then decrypted through paillier homomorphic decryption technique to get the final results. These final results are then transferred to doctor, patients and hospitals.

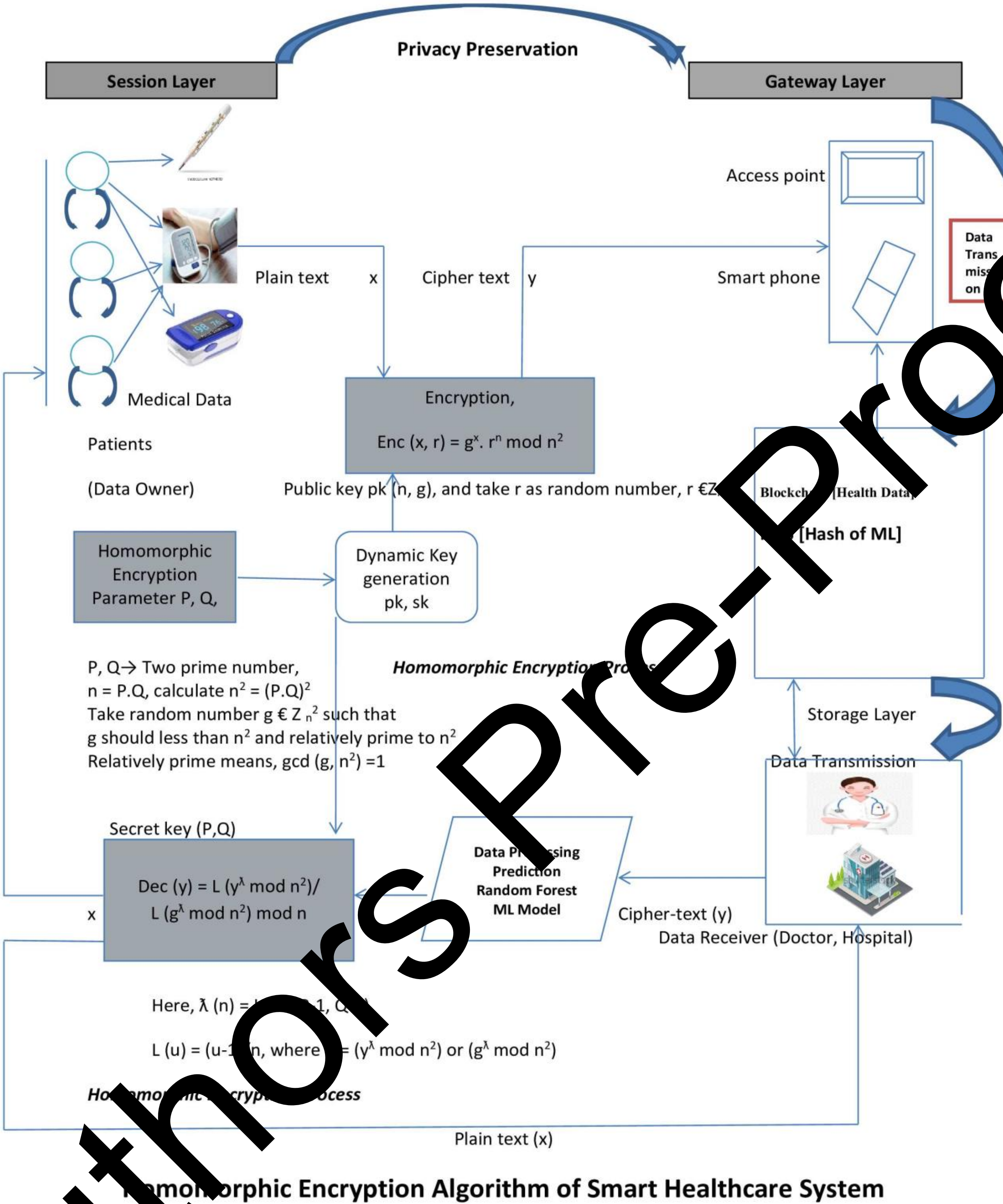


Fig 4. Proposed Architecture of Paillier Homomorphic Encryption and Blockchain Technology based Secured Healthcare System

IV. DESIGN AND ANALYSIS

Now a day, Internet of Medical Things (IoMT) an exceptional type of IoT network which is implemented in SHS where privacy and security are the key concern to be handled [27]. The integration of the IoT especially SHS network with traditional healthcare systems has improved quality of healthcare services [28]. However, the wearable devices and sensors used in Healthcare System (HS) continuously monitor and transmit data to the nearby devices or servers using an unsecured open channel [29]. With the help of block chain and ML technique, healthcare fifth generation system is developed into smart healthcare system [30]. The basic objective of smart healthcare is to reduce patient stress and healthcare cost. Machine learning is a technique where a central server is used in SHS network. In our research, we will try to develop a secure and protected data transmission framework for a smart healthcare system which will be designed using block chain technology and machine learning method using cryptographic algorithm at different level where an intrusion detection system will detect any unauthorized activities in the healthcare network. Doctors can monitor patient's condition continuously using various medical sensors which will be attached with patient and will be connected to the SHS network. These sensor will collect patients medical information like temperature, heart bit, BP etc. from time to time and can transfer these information to doctor so that doctor can take necessary medical step for the patients through this smart healthcare system. In this proposed framework, several numbers of medical organizations are connected through IoMT network which is directly connected to a global network so that all the clinical data can be shared very securely throughout the IoMT based smart healthcare network.

Numerous networking applications have been transformed by innovative IoT solutions made possible by the quick advancements in micro-computing, mini-hardware manufacturing and machine-to-machine (M2M) communications [31]. One of the applications that IoT has revolutionized is healthcare systems [32]. To this end, an IoT branch called IoMT systems has been introduced and is used to design SHS. Patients with chronic illnesses can be remotely monitored thanks to SHS. As such, it can offer patients prompt diagnostics that, in an emergency, may save their lives. Nonetheless, one of the biggest obstacles to these vital systems' widespread use is security. Modern methods for safeguarding data from smart healthcare systems during collection, transmission, and storage are discussed in this article. We provide a thorough assessment of all possible network and physical threats on SHS systems[33]. Our research shows that most security methods do not take different kinds of assaults into account. Thus, we provide a security architecture that incorporates a number of security measures. The majority of known attacks may be mitigated by the framework, which also covers SHS security standards. In the proposed SHS based layered architecture of smart healthcare system, there are four basic layers: Session layer, gateway layer, cloud layer, and visualization layer which jointly perform the complete functionalities of smart healthcare system.

Proposed SHS Layered Architecture

Figure 4 represents the layered architecture of smart healthcare system where four basic layers are being used: session, gateway, storage, and visualization layer.

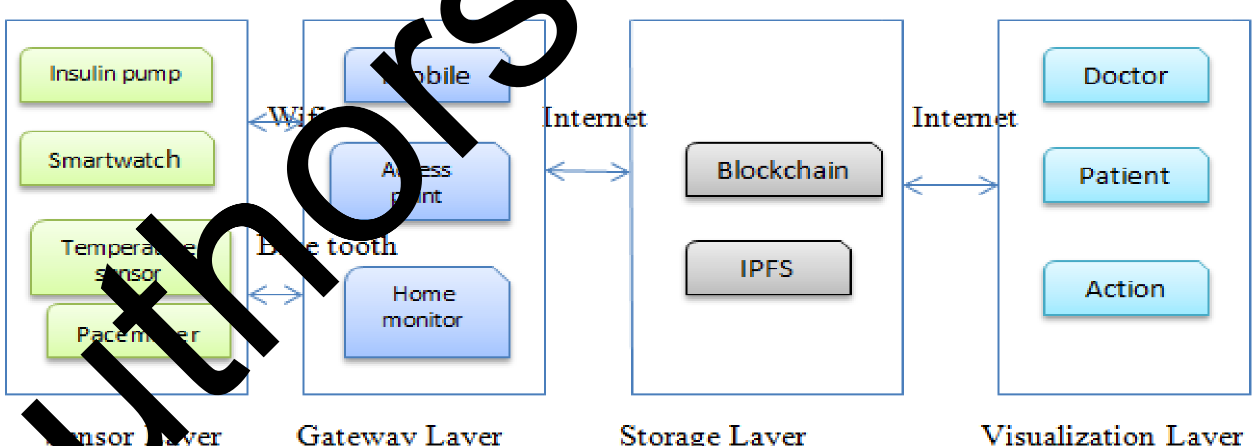


Fig 5. Layered Architecture of the Proposed Smart Healthcare System

1. Sensor Layer: A collection of tiny sensors that are worn or implanted and gather the patient's biometrics make up this layer. The second layer receives the data using wireless protocols like Wi-Fi, Bluetooth, or the MedRadio frequency (RF) band.

2. Gateway Layer: Owing to IoMT sensors' limited processing and storage capacity, data are sent to the gateway layer—the second layer—without being processed. This layer's gadgets, which are typically more potent than sensors, can be the patient's smartphone or a specific access point (AP). Some pre-processing tasks, such validation, temporary data storage, and basic AI-based analysis, are within their capabilities. Furthermore, the sensor data is transmitted via the Internet to the cloud.

3. Cloud Layer: Getting the data from the gateway for safe access, analyzing, and storing are the responsibility of the cloud layer. In order to identify any changes in the patient's health, the analysis may involve analyzing the data and providing it to the doctors or patients for additional action. IDs and keys for different system nodes are generated by the key generation server (KGS). From this layer, remote management and control of the sensor access is possible.

4. Visualization/Action Layer: Patients and doctors can track each other's health with the data in this layer. The doctor's recommendations for the patient's course of treatment are also included in this layer. Medication prescriptions and dose adjustments are two examples of activities.

Proposed Work Flow

Step-1. Registration and Authentication Scheme: 1. Registration 2. Login 3. Verification

Step-2. Privacy Preservation Process (Paillier Homomorphic Encryption)

Step-3. Healthcare Data and ML Model Storing (Blockchain Technology and IPFS)

Step-4. Healthcare Data Prediction (Random Forest ML Algorithm)

Step-5. Data Assessment/Visualization

Step-1. Registration and Authentication Scheme

Prior to uploading health data to the blockchain and SHS, patients, physicians, and hospitals must first register with the SHS. The three phases that make up the authentication process are registration, login, and verification. During user registration, the user selects the identity IDi, password PWDi, and random number ki. IDi's presence in the SHS database is verified by the database. In the event that it is not, the SHS records that data and alerts the user that the device ID has already been registered. The user provides the SHS database with their login credentials (LC) information during login. Subsequently, the SHS confirms that the user's information and ID are included in the dataset. The SHS permits users to transmit information if it is accessible. In addition to the SHS registration site, the management registration window is available. In that window, the SHS-related information is input. Following verification, the health service accesses the SHS database if the login information is included in the dataset.

Step-2. Privacy Preservation Process (Paillier Homomorphic Encryption)

The privacy preservation process is maintained between session layer and gateway layer with implementing paillier homomorphic encryption algorithm. This implies that after collecting the patient's medical data, when this sensitive health data is transmitted from medical sensor to another access point i.e computer, smart phone, laptop which are owned by doctor, diagnostic center, and hospital, Paillier encryption algorithm is implemented for data privacy and security purpose. The prime objective is to make secure the health data while transmitting throughout one point to another point of SHS.

The Paillier encryption algorithm has the following four (04) phases:

Assume that, taking the patient health data as plain text x, and after encryption, the cipher text y,

1. Key Generation:

Consider homomorphic encryption parameter M and N

M, N→ Two prime number,

$r = M.N$, Determine $r^2 = (P.Q)^2$

Take random number $g \in Z_{n^2}$ such that g should less than r^2 and relatively prime to r^2,

Relatively prime means, $gcd(g, r^2) = 1$,

Here, $\partial(n) = LCM(M-1, N-1)$

$L(u) = (u-1)/r$, where $u = (y\partial \bmod r^2)$ or $(g\partial \bmod r^2)$

2. Public key and private key:

Public key, PK (r, g), and take t as random number, $r \in Z_n$,

Secret key, SK (M, N)

3. Encryption process:

Encryption, $Enc(x) = y = g^x . r^n \bmod r^2$

4. Decryption process:

Dec (y) = $L(y\partial \bmod r^2) / L(g\partial \bmod r^2) \bmod r$

The homomorphic property of Paillier encryption permits us to implement addition on cipher texts. If we have two cipher texts c_1 and c_2 encrypting plaintexts m_1 and m_2 respectively. Then multiplying $(c_1 \times c_2) \bmod n^2$ decrypts to $(m_1 + m_2) \bmod n$

Besides being homomorphic for addition, Paillier encryption also facilitates "homomorphic multiplication." That being said, this necessitates a further step known as "homomorphic re-encryption," which is increasing a cipher text to an exponent that depends on the plaintext. With encrypted data, this feature makes more intricate procedures possible.

Step-3. Healthcare Data and ML Model Storing (Blockchain Technology and IPFS)

In this step, the healthcare dataset which is received by the gateway layer is reserved in the blockchain to guarantee data privacy and security. The blockchain may be used to store healthcare data as transactions. A particular medical record or data item may be represented by each transaction. To address data security concerns, the trained ML model was not directly stored in the blockchain. Instead, the model was stored in IPFS, a decentralized storage system, ensuring data integrity and availability. A hash of the model stored in IPFS which is then recorded on the blockchain, providing a secure and immutable reference to the model.

Machine learning (ML) models are stored in smart healthcare systems using IPFS (Inter Planetary File System), mainly because of its distributed and decentralised architecture. ML models stored on IPFS are more available and less dependent on a single point of failure as they may be accessed and used by several nodes within the network without the need for a central server. To further aid in confirming the integrity of the models, IPFS furthermore offers content addressing, guaranteeing that every ML model has a distinct hash derived from its content. The ideas of smart healthcare systems, which emphasise scalability and security through the distribution and accessibility of data and resources throughout a network, are in line with this decentralised approach.

Blockchain lacks administrative authorization to alter or remove data; it is a write-only data structure. It is used by healthcare facilities to address problems with healthcare data security. Health records are secured inside the cloud architecture by being shared, encrypted, and held by several parties. A block has a record allocated to it, and it is logically connected to the previous block. As a result, records have a connection and are not compressed; instead, the blocks contain their timestamps, making network transaction verification simple. Every block in the chain has a header and a list of legal connections. The header comprises a number of elements related to the network's settings (like mining parameters) and the data structure's integrity (like the timestamp). User submissions of transactions, which alter the network's status, are also possible from nodes. Blockchain refers to the concept whereby connected blocks create a full chain. Building elements like as distributed ledgers, consensus, smart contracts, and data blocks enable the creation of a blockchain network.

Step-4. Healthcare Data Prediction (Random Forest ML Algorithm):

In recent years, the integration of advanced technologies such as machine learning (ML), Inter Planetary File System (IPFS), and blockchain has revolutionized various industries, including healthcare. This report explores the utilization of these technologies to predict healthcare data, ensuring both accuracy and security. The primary objective is to develop a robust system capable of predicting healthcare outcomes using ML models, while also addressing concerns regarding data security through the implementation of IPFS and blockchain technology.

Random Forest Algorithm

Step 1: From the training set, choose M data points randomly.

Step 2: Create the decision trees linked to the chosen data points (subsets).

Step 3: Select the number X for the decision trees you wish to construct.

Step 4: Continue Steps 1 and 2.

Step 5: Localize each decision tree's predictions for the newly data points, and then assign them to the category that receives the maximum of the votes.

The prediction of a Random Forest model may be expressed mathematically as follows:

A. ***Classification****: y^=mode (f1(z), f2(z),...,fr(z))*

Where y^ is the predicted class,

fi(z) is the prediction of the i-th decision tree, and

r is the total number of trees in the forest.

B. ***Regression****: y^=1/r* $\sum_{i=1}^{r} fi(z)$

Where y^ is the predicted target value,

fi(x) is the prediction of the i-th decision tree, and

n is the total number of trees in the forest.

Performance of Random Forest (RF) Algorithm:

The performance of this proposed procedure can be calculated by using some parameters like precision, recall, F1 score, sensitivity, and confusion matrix. The following is a typical confusion matrix for a binary classification issue with classes "Positive" (P) and "Negative" (N):

	Predicted Positive (P')	Predicted Negative (N')
Actual Positive (P)	True Positive (TP)	False Negative (FN)
Actual Negative (N)	False Positive (FP)	True Negative (TN)

The following three steps can be followed to calculate and evaluate all the performance metric parameters of the proposed RF algorithms:

Step 1: First, build the confusion matrix using the real labels and our Random Forest forecasts.

Step 2: Utilizing the confusion matrix, compute TP, FP, TN, and FN.

Step 3: Calculate Accuracy. This provides an overall indicator of the classifier's frequency of accurate predictions:

- ✓ *Accuracy = Correction Prediction Quantity / Total Prediction Quantity*
  = (TP + TN) / (TP + FP + FN + TN)

Step 4: Determine Precision, which provides the percentage of positive identifications that were truly accurate.

- ✓ *Precision = TP / (TP + FP)*

Step 5: Determine Recall (also known as Sensitivity), a metric that expresses how many positive class instances were accurately detected:

- ✓ *Recall = TP / (TP + FN)*

Step 6: Calculate the F1 Score, which attacks a stability between precision and recall by taking the harmonic mean of the two:

- ✓ *F1 Score = 2 * Precision * Recall / (Precision + Recall)*

Since Random Forest can handle huge and complicated datasets and is strong against overfitting, it is the favoured method for healthcare data prediction in smart healthcare systems. In the medical field, where precise forecasts are essential for favourable patient outcomes, Random Forest's feature significance analysis offers valuable perspectives on significant variables. It is an appropriate option for healthcare applications where interpretability and speed are critical due to its capacity to handle missing values without imputation, parallel processing for quicker performance, and ensemble learning for increased prediction reliability. In short, Random Forest offers resilience, accuracy, and generalisation capabilities by combining the predictions of several decision trees to provide precise predictions on healthcare data.

Procedure:

1. Machine Learning Model Training (Random Forest):

A. The ML model, specifically a RF algorithm, was selected for its capability to manage complex datasets and offer exact predictions in healthcare applications.

B. Python programming language was utilized for implementing the Random Forest algorithm, leveraging libraries such as scikit-learn.

2. Frontend and Backend Development:

A. The frontend of the application was developed using web development technologies, providing an intuitive user interface for interacting with the system.

B. Flask, a Python web framework, was employed for building the backend of the application, facilitating communication between the frontend and the ML model.

3. User Interaction and Prediction:

A. Users have the option to update the ML model through the frontend interface.

B. Upon receiving input values from the user, the backend processes the data using the trained ML model to predict healthcare outcomes.

C. The results of the prediction are then displayed on the frontend interface for the user.

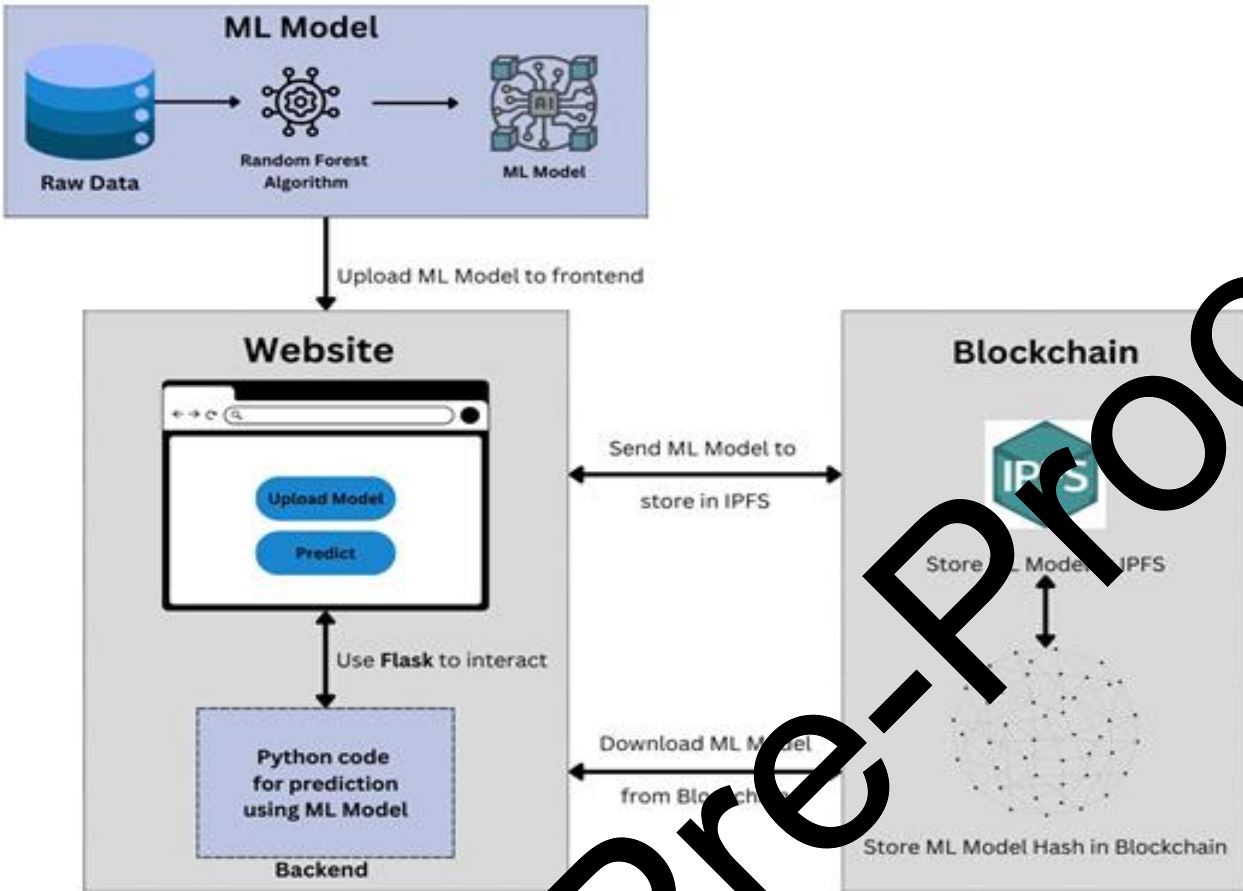


Fig 6. Blockchain and IPFS based Healthcare Data Prediction using ML (Random Forest Model) of Smart Healthcare System

Step-5. Data Assessment/Visualization:

In this last step, the resulting data can be accessed or visualized by various access point devices which are owned by patient, doctor, and hospital. After performing the healthcare data prediction process by using Random Forest ML model, the decision which is in the encrypted form, will be decrypted again through Paillier Homomorphic algorithm. Then the final result which will be then in the plain text form will be transferred to patient, doctor, and hospital. Finally, health service centres evaluate the patient's condition using the decrypted data.

V. RESULTS AND DISCUSSION

In this Paillier homomorphic encryption and blockchain based smart healthcare system, the data security level after encryption and decryption is 100%. The accuracy level of healthcare data prediction of heart disease dataset is 90.16% by using Random Forest ML model. Let us focus on the following table-1, where different data accuracy level of healthcare data prediction using various ML algorithms is represented. Here, data is taken from existing research works on healthcare system.

Table 1: Comparative Evolution of Our Research Findings with Previous Research

SL No.	ML Algorithm	Accuracy (%)
1.	Logistic Regression	85.25
2.	Naïve Bayes	85.25
3.	Support Vector Machine	81.97
4.	K-Nearest Neighbors	67.21
5.	Decision Tree	81.97
6.	XG Boost	85.25
7.	Neural Network	80.33

9.	Random Forest [Proposed]	90.16

Following is the heart disease dataset used in our proposed smart healthcare system.

```
dataset = pd.read_csv("newheart.csv")
```

```
dataset
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	[illegible]
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

The following figure 7 represents the graphical presentation of various ML model based healthcare data prediction accuracy level in percentage. From this figure 7, it is clear that out of various ML algorithms, Random Forest method has the maximum accuracy level 90.16% as compared to other ML models. This figure is generated from the above table 1.

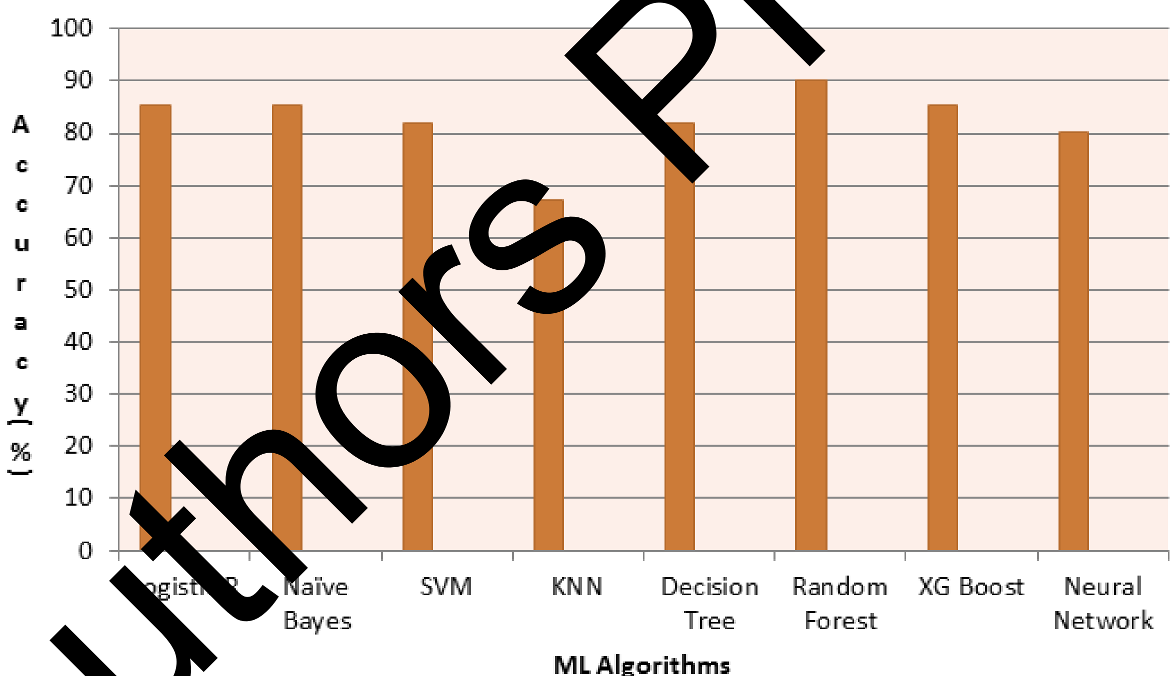


Fig 7. Graphical Representation of Accuracy Level (%) of Healthcare Data Prediction using Various ML Model in SHS

Different data security level (in percentage) using various cryptographic algorithms is represented in the following table 2. Here, data is taken from existing research works on healthcare system.

Table-2: Different Security Level of Data using Cryptographic Algorithms

SL No.	Cryptographic Algorithm	Security (%)
1.	ECC	87
2.	RSA	84
3.	DES	83
4.	**Paillier Homomorphic Encryption [Proposed]**	**100**

The following figure 8 represents the graphical presentation of various Cryptographic algorithm based data security level in percentage. From this figure 8, it is observed that out of various Cryptographic algorithms, our proposed Paillier Homomorphic Encryption method has the highest level of security 100% as compared to other cryptographic algorithms. This figure is generated from the above table 2.

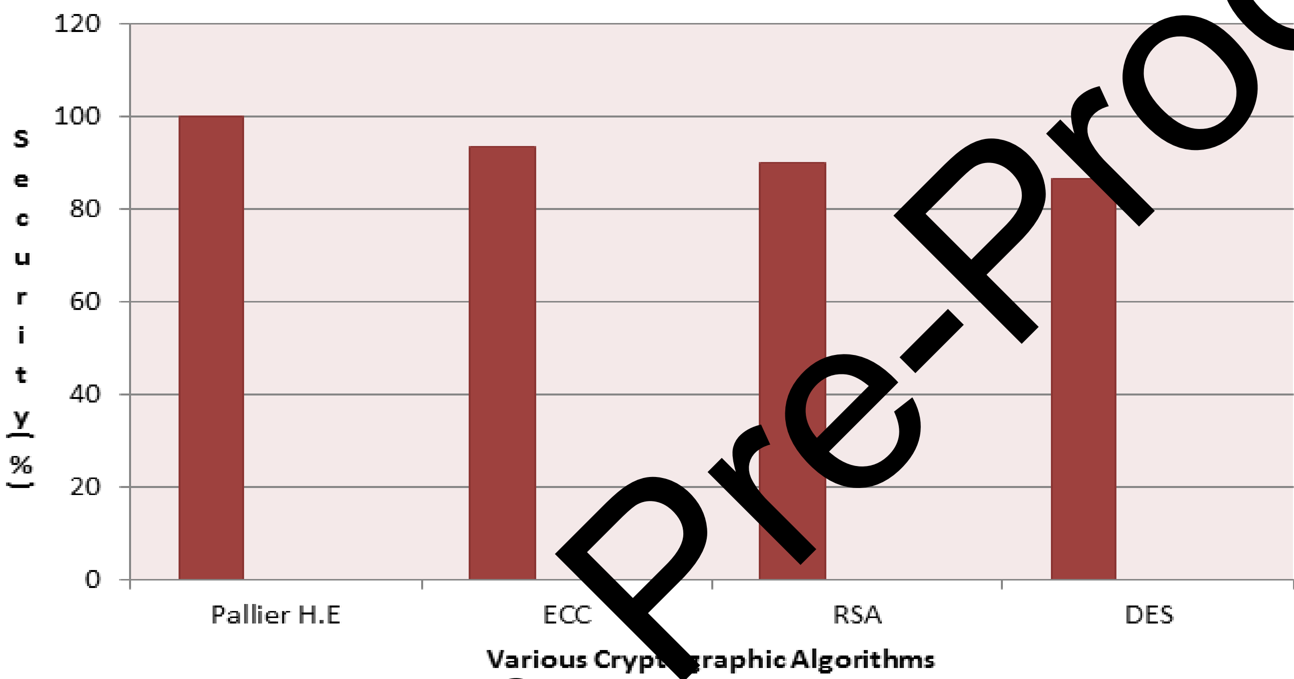


Fig 8. Graphical Representation of Accuracy Level (%) of Healthcare Data Prediction using Various ML Model in SHS

VI. FUTURE WORK

To develop the efficacy, safety, and scalability of Paillier homomorphic encryption, blockchain technology, and Random Forest machine learning in smart healthcare systems, a number of directions might be investigated in future study. The machine learning model used to predict cardiac disease might be improved in numerous ways. For example, by investigating different algorithms or ensemble techniques, accuracy and resilience could be increased even more. Research may also concentrate on creating encryption and decryption methods that are more effective in order to minimise computational cost and preserve data security. Future research should focus on integrating other data sources, such lifestyle or genetic data, to improve the system's forecasting skills.

Additionally, real-world deployment and assessment studies examining the system's influence on clinical decision-making and patient outcomes may offer important new perspectives on the usefulness and efficacy of the system in healthcare environments. In order to overcome existing shortcomings and further progress the subject of smart healthcare systems, future research should generally focus on improving and expanding upon the suggested system.

VII. CONCLUSION

In order to secure health data and anticipate heart disease, we in this study suggested a unique method for developing and evaluating a smart healthcare system that combines blockchain technology, the Random Forest machine learning algorithm, and the Paillier homomorphic encryption algorithm. Predictive analytics and safe health data exchange in the healthcare industry are difficulties that are addressed by the combination of these technologies.

Health data is kept private during transmission and storage thanks to the Paillier homomorphic encryption technique, and blockchain technology offers a visible and safe ledger for tracking medical transactions. Heart disease prediction uses the Random Forest algorithm, which analyses encrypted data.

Compared to conventional techniques, our study shows that the integrated strategy greatly improves health data security and prediction accuracy. Sensitive data is kept private thanks to the Paillier homomorphic encryption technique, while blockchain technology offers an unchangeable ledger of data access and exchange. Additionally, the Random Forest algorithm delivers high prediction accuracy for cardiac illness, giving medical practitioners important information to help them make judgements. Simulation tests were used to assess the system's performance and validate its efficacy in terms of data security, prediction accuracy, and computing proficiency. To sum up, the suggested smart healthcare system provides a thorough approach to guaranteeing the security of patient data and enhancing predictive analytics in the medical field. Its use of the Random Forest algorithm, blockchain technology, and Paillier homomorphic encryption shows how it may improve healthcare services and advance the field of smart healthcare systems. Subsequent research endeavours may focus on refining the system's functionality and broadening its use in various healthcare fields.

References:

1. Budida, D.A.M. and Mangrulkar, R.S., 2017, March. Design and implementation of smart HealthCare system using IoT. In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) (pp. 1-7). Ieee.
2. Dwivedi, R., Mehrotra, D. and Chandra, S., 2022. Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: A systematic review. Journal of oral biology and craniofacial research, 12(2), pp.302-318.
3. Li, W., Chai, Y., Khan, F., Jan, S.R.U., Verma, S., Menon, V.G., Kavita, F. and Li, X., 2021. A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mobile networks and applications, 26, pp.234-252.
4. An, B.W., Shin, J.H., Kim, S.Y., Kim, J., Ji, S., Park, J., Lee, Y., Jang, J., Park, Y.G., Cho, E. and Jo, S., 2017. Smart sensor systems for wearable electronic devices. Polymers, 9(8), p.303.
5. Cooper, R.A., Dicianno, B.E., Brewer, B., LoPresti, E., Ding, D., Simpson, R., Grindle, G. and Wang, H., 2008. A perspective on intelligent devices and environments in medical rehabilitation. Medical Engineering & Physics, 30(10), pp.1387-1398.
6. Razdan, S. and Sharma, S., 2022. Internet of medical things (IoMT): Overview, emerging technologies, and case studies. IETE technical review, 39(4), pp.775-788.
7. Kushwaha, P.K. and Kumaresan, M., 2021, November. Machine learning algorithm in healthcare system: A Review. In 2021 International Conference on Technological Advancements and Innovations (ICTAI) (pp. 478-481). IEEE.
8. Qureshi, K.N., Din, S., Jeon, G. and Piccialli, F., 2020. An accurate and dynamic predictive model for a smart M-Health system using machine learning. Information Sciences, 538, pp.486-502.
9. Rahmani, A.M., Yousefpoor, E., Yousefpoor, M.S., Mehmood, Z., Haider, A., Hosseinzadeh, M. and Ali Naqvi, R., 2021. Machine learning (ML) in medicine: Review, applications, and challenges. Mathematics, 9(22), p.2970.
10. D'souza, K.J. and Ansari, Z., 2018, November. Big data science in building medical data classifier using Naïve Bayes model. In 2018 IEEE international conference on cloud computing in emerging markets (CCEM) (pp. 76-80). IEEE.
11. Harimoorthy, K. and Thangavelu, M., 2021. RETRACTED ARTICLE: Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. Journal of Ambient Intelligence and Humanized Computing, 12(3), pp.3715-3723.
12. Al-Turjman, F., Nawaz, M.H. and Ulusar, U.D., 2020. Intelligence in the Internet of Medical Things era: A systematic review of current and future trends. Computer Communications, 150, pp.644-660.
13. Nidhya, R., Kumar, M., Maheswar, R. and Pavithra, D., 2022. Security and privacy issues in smart healthcare system using internet of things. IoT-Enabled Smart Healthcare Systems, Services and Applications, pp.63-85.
14. [illegible], J., [illegible], Q., Ji, Y., Xu, M. and Xue, R., 2022. Secure medical data management with privacy-preservation and authentication properties in smart healthcare system. Computer Networks, 212, p.109013.
15. Singh, A., Sukhija, N., Sharma, A., Gupta, M. and Aggarwal, P.K., 2021. Security and Privacy Requirements for IoMT-Based Smart Healthcare System: Challenges, Solutions, and Future Scope. In Big Data Analysis for Green Computing (pp. 17-37). CRC Press.
16. Awotunde, J.B., Jimoh, R.G., Folorunso, S.O., Adeniyi, E.A., Abiodun, K.M. and Banjo, O.O., 2021. Privacy and security concerns in IoT-based healthcare systems. In The fusion of internet of things, artificial intelligence, and cloud computing in health care (pp. 105-134). Cham: Springer International Publishing.
17. Li, W., Chai, Y., Khan, F., Jan, S.R.U., Verma, S., Menon, V.G., Kavita, F. and Li, X., 2021. A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mobile networks and applications, 26, pp.234-252.
18. Motwani, A., Shukla, P.K. and Pawar, M., 2022. Ubiquitous and smart healthcare monitoring frameworks based on machine learning: A comprehensive review. Artificial Intelligence in Medicine, 134, p.102431.
19. Zamzami, I.F., Pathoee, K., Gupta, B.B., Mishra, A., Rawat, D. and Alhalabi, W., 2022. Machine learning algorithms for smart and intelligent healthcare system in Society 5.0. International Journal of Intelligent Systems, 37(12), pp.11742-11763.
20. Saif, S., Jana, M. and Biswas, S., 2021. Recent trends in IoT–based smart healthcare applying ml and dl. Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3, pp.785-797.
21. Nasr, M., Islam, M.M., Shehata, S., Karray, F. and Quintana, Y., 2021. Smart healthcare in the age of AI: recent advances, challenges, and future prospects. IEEE access, 9, pp.145248-145270.

22. Chatrati, S.P., Hossain, G., Goyal, A., Bhan, A., Bhattacharya, S., Gaurav, D. and Tiwari, S.M., 2022. Smart home health monitoring system for predicting type 2 diabetes and hypertension. Journal of King Saud University-Computer and Information Sciences, 34(3), pp.862-870.
23. Ramesh, J., Aburukba, R. and Sagahyroon, A., 2021. A remote healthcare monitoring framework for diabetes prediction using machine learning. Healthcare Technology Letters, 8(3), pp.45-57.
24. Ghazal, T.M., Rehman, A.U., Saleem, M., Ahmad, M., Ahmad, S. and Mehmood, F., 2022, February. Intelligent model to predict early liver disease using machine learning technique. In 2022 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-5). IEEE.
25. Raeesi Vanani, I. and Amirhosseini, M., 2021. IoT-based diseases prediction and diagnosis system for healthcare. Internet of Things for Healthcare Technologies, pp.21-48.
26. Ray, A. and Chaudhuri, A.K., 2021. Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development. Machine Learning with Applications, 3, p.100011.
27. Ganesan, M. and Sivakumar, N., 2019, March. IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. In 2019 IEEE international conference on system, computation, automation and networking (ICSCAN) (pp. 1-5). IEEE.
28. Zamzami, I.F., Pathoee, K., Gupta, B.B., Mishra, A., Rawat, D. and Alhalabi, W., 2022. Machine learning algorithms for smart and intelligent healthcare system in Society 5.0. International Journal of Intelligent Systems, 37(12), pp.11742-11763.
29. Lou, Z., Wang, L., Jiang, K., Wei, Z. and Shen, G., 2020. Reviews of wearable healthcare systems: Materials, devices and system integration. Materials Science and Engineering: R: Reports, 140, p.100523.
30. Mittal, P., 2023. Fusion of Machine Learning and Blockchain Techniques in IoT-based Smart Healthcare Systems. In Deep Learning for Healthcare Decision Making (pp. 245-266). River Publishers.
31. Praveenchander, J., Vetrithangam, D., Kaliappan, S, Karthick, M, Pegada, N. R., Patil, P. P. & Umar, S (2022). IoT-Based Harmful Toxic Gases Monitoring and Fault Detection on the Sensor Dataset Using Deep Learning Techniques, Scientific Programming, 2022(1), 7516328.
32. Vetrithangam, D., Senthilkumar, V., Kumar, A. R., Naresh, P, & Sharma, M. (2022). Coronary Artery Disease Prediction Based on Optimal Feature Selection u Improved Artificial Neural Network with Meta-Heuristic Algorithm. Journal of Theoretical and Applied Information Technology, 24.
33. Raza, A., Ali, M., Ehsan, M. K., & Sodhro, A. H. (2023). Spectrum Evaluation in CR-Based Smart Healthcare Systems Using Optimizable Tree Machine Learning Approach. Sensors, 23([illegible]), [illegible].
34. Azeez, N. A., & Van der Vyver, C. (2019). Security and privacy issues in e-health cloud-based system: A comprehensive content analysis. Egyptian Informatics Journal, 20(2), 97-108.