Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Secure Heart Disease Classification System Based on Three Pass Protocol and Machine Learning

Randa shaker Abd-Alhussain*[1], Hadab Khalid Obayes[2], Farah Al-Shareefi[3]

[1]College of science for women, University of Babylon,Babylon,51002, Iraq
[2] College of education for humanities studies, University of Babylon,Babylon, 51002, Iraq
[3] College of science for women, University of Babylon,Babylon, 51002, Iraq

*Corresponding Author: Randa Shaker

DOI: https://doi.org/10.52866/ijcsm.2023.02.02.003
Received November 2023 ; Accepted Jenury 2023 ; Available online Februry 2023

ABSTRACT: Heart disease is one of the worst life-threatening conditions. Correct and early diagnosis of this disease is crucial for saving patients' life and avoiding other complications. On the other hand, keeping the patient's data, diagnosis process, and treatment plan secured is equally important to the defacto medical procedure. This research proposes a system that is consisting of two phases: security provision and patients' condition diagnosis. Typically, the first phase exercises a security protocol, called three-pass protocol, to ensure that the people who can access the patient's information are authorized. In order to obtain a high accuracy level in the diagnosis process, artificial intelligence with machine learning methods are employed in the later phase. The proposed system relies on a data set which includes a number of vital indicators, by which the patient's status can be classified as having heart disease or not. The K-Nearest Neighbor (KNN) algorithm and the random forest tree algorithm are applied to carry out the classification task. The accuracy scale results reveals that the random forest tree algorithm (99%) gave higher accuracy than KNN (97%).

Keywords: Hear t Disease, KNN, Random Forest, Three Pass Protocol, Security

1. INTRODUCTION

Globally, heart diseases are the major cause of human morbidity and mortality. In order to save the patients' life and prevent further deterioration in their health, an early and accurate diagnosis is undeniably vital. Currently, the leading technologies supported by Artificial Intelligence (AI) are utilized in diagnosing heart diseases accurately[1]. One of the AI applied algorithms is machine learning. A range of strategies are used in machine learning to assist in enhance making decisions in the diagnosis process according to medical data[2][3]. However, choosing a machine learning technique with the most precise performance is still under scrutiny. In addition, any diagnosis process needs to determine a very cautiously selected set of characteristics representing the disease conditions.

Along with the diagnosis related concerns, more attention should be paid to protect the patients' bio-medical data from any unauthorized access by hackers and criminals. In order to ensure the security of data, encryption methods can be used. Cryptography often uses mathematical techniques to carry out encryption and decryption on sent data, also known as messages[4].Cryptography is classified into two types: symmetric-key cryptography, and public-key cryptography [5]**.** A common and secure symmetric-key technique for encrypting important data is called Advanced Encryption Standard (AES). Key distribution, however, is one of the problems connected with this technique [6].In other words, the AES assumes that the key is securely dispensed among the participants and that there is an implicit trust established between them, which is not true in all cases. Fortunately, the cryptography algorithms are used by protocols, called cryptographic protocols. These protocols achieve one or more security services, such as: key distribution, authentication, secrecy, etc. [7]. One of these protocols is Three-Pass (TP) protocol, which is developed to exchange secret messages between the sender and the receiver without a prior key exchange between them[8].

According to the above debate, this paper focuses on executing the TP protocol to tackle the AES's issue and to reinforce the confidentiality of patient's information. Furthermore, this research shows how to build an effective and safe healthcare system while taking advantage of the dataset on heart disorders already accessible, in order to create an algorithm that aids in the diagnosis and classification of heart patients using a set of clinical markers. Compared to previous relevant research and mainly for the need to improve the performance accuracy. essentially, we analyze the

performance of two machine learning techniques in classification action, namely, the K-Nearest Neighbor (KNN)classifier [9] and the Random Forest model[10].

The major contributions of this paper include:

1) designing an automated system to diagnose heart diseases with a dual aim of improving the disease classification's accuracy using machine learning-based techniques, and securing patient data from intended alteration or hacking;
2) implementing the Three-Pass protocol to securely distribute a key and confidential information between participants without any prior knowledge between them;
3) evaluating the effectiveness of the two classification machine learning methods utilizing a variety of performance metrics.

The rest of the paper is organized as follows. Section 2 covers the pertinent literature. Section 3 recapitulates the main methods and techniques utilized in this paper. A detailed description of the proposed system is documented in Section 4. Section 5 discusses our results and findings, while Section 6 recaps this paper.

2. RELATED WORK

As our work comprises of identifying the presence of heart problems and preserving the patient's data secure, this section reviews the literatures that have been accomplished in the similar fields.

A smart heart-disease prediction system has been developed by the authors of [11]. They have utilized the Naive Bayesian (NB) classification method and the Advanced Encryption Standard (AES) algorithm in their system. They have prepared the medical profiles, which contain the following attributes: patient's age, blood pressure, cholesterol, sex, blood sugar, and other characteristics, to be inputs to the NB. The developed system consists of several steps, including dataset gathering, user enrollment and login, classification using NB, prediction, and securing the transferred data using the AES algorithm. Their technique showed that the NB achieves an accuracy of 89.77%, even when the attributes are minimized. Furthermore, comparing to the parallel homomorphic encryption algorithm, AES produces higher security performance results.

Different from our work, the researchers of [12] have redefined the Three Pass protocol via utilizing an equation introduced by Schrödinger that is served a physical system called a boson system. In addition, a comparative analysis of two well-established cryptographic techniques, namely, RSA and DES, has been presented in [13]. This analysis has been carried out using a type of medical database, called horizontally partitioned database. According to its experimental findings, the DES takes less time but is less secure, and the RSA takes more time but is more secure. As a result, a hybrid solution combining these two techniques has been developed to be faster than DES and RSA while also being more secure.

In[14], an automatic tool is engineered to aid clinicians in the early diagnosis of heart issues. The various machine learning techniques based on a quick analysis of heart disease diagnosis are presented in this research. First, heart disease is predicted using NB with a weighted method. Second, an automatic analysis of the localization of cardiac ischemia in accordance with the characteristics of frequency, time, and information theory. In this step, the two most effective classifiers: support vector machine (SVM) and XGBoost, are chosen for the classification. Third, an upgraded SVM based on the duality optimization strategy that has been also analyzed to identify heart failure. Finally, an efficient Heart Disease Prediction Model (HDPM) is used for a Clinical Decision Support System (CDSS).The analysis results have shown that the XGBoost algorithm for detecting heart disease has high accuracy, precision, recall, and F1-measure values. In contrast, the NB with a weighted approach only has low accuracy, and the SVM with duality optimization has low precision, recall, and F1-measure values.

The authors of [15] addressed the issue of security and privacy guarantees because illness risk prediction cannot continue to develop without them. Basically, they were developed what is referred to as EPDP, an Effective and Privacy-preserving Disease risk Prediction system for telehealth. It accomplishes two stages of illness risk prediction, namely disease model training and disease prediction, while guaranteeing the preservation of privacy. In the phase of illness model training, the symptom set of each disease is extracted using a combination of super-increasing series and homomorphic cryptographic technique. In the prediction phase, the bloom filter approach is introduced to compute the prediction results. They demonstrated how well suited their technology is to medical emergencies.

The researchers of [16] have created a method for analyzing a patient's ECG data to predict what kind of arrhythmia is present. For achieving the prediction action, machine learning techniques, such as Support Vector Machine, K-Nearest Neighbors, Convolution Neural Network, and Random Forest, are carried out in this method. In addition, this method also includes preventing fraud and data breaches by performing blockchain technology. Their obtained results have shown that the convolution neural network gives the highest accuracy among the used machine learning techniques. However, their results seemingly do not show how the blockchain method could reduce cybercrime risks, such as data fraud.

Artificial intelligence based computer-aided diagnosis approach has been designed in [17] to lower death rates by assisting specialists in making decisions and enabling early. This method consists of three modules: an electrocardiogram processor module that is dedicated for analyzing various waveforms, a machine learning based

prediction module which uses clinical and patient information to predicate heart disease, finally a deep learning based classification module that depends on 12-lead ECG for classifying 18 cardiac conditions.
The authors of [18] have applied eleven machine learning classifiers, such as Gradient Boosted Tree, Multilayer Perceptron, Random Forest, etc., for predicting the occurrence of heart problems. They have used UCI repository to obtain the dataset. Among the applied classifiers, they found that the Random Forest classifier yields the highest accuracy of 96.28%.
A computational model for detecting heart abnormality is developed by the authors of[19]. This model incorporates a suggested regularizer with artificial neural network. A regularizer, in general, refers to the process of avoiding data memorization throughout the training stage. The suggested regularizerh as penalized attribute coefficients from having high values in the space of the weight matrices in order to perform well on the test data. The obtained accuracy with this model was 96.30%.

3. MATERIALS AND METHODS

This section briefly surveys the main techniques and materials that are employed in our work.

3.1 CRYPTOGRAPHIC PROTOCOLS

Cryptographic or authentication protocols are pre-described and distributed procedures that utilize the cryptographic algorithms to attain security service(s), like data confidentiality, in a suspicious environment [20]. Formally, a protocol is a finite series of message exchanging steps, such that every step has the following form:

$$i. A_i \rightarrow B_i: M_i \qquad (1)$$

Where $1 \leq i \leq n$, is the i[th] protocol's step, n is a total number of the protocol's steps, A_i is the sender of step i, B_i is the receiver of step i, and M_i is the message of that step. This form denotes that M is sent from participant A to participant B. There are many examples of the cryptographic protocols, but we will only state the three pass protocol below, which has been utilized in this paper. Furthermore, each protocol uses different encryption methods in its work. From these methods, only the AES will be described in the below subsection.

3.1.1 THREE – PASS PROTOCOL

The Three-Pass (TP) protocol is developed to pass a private message over an insecure communication network, without the requirement of sharing or distributing any secret key [8]. This protocol, as its name suggests, is designed to exchange three messages, which are:

$$1. A \rightarrow B: \{m\}_{K(A)}$$

$$2. B \rightarrow A: \{m\}_{K(A)}\}_{K(B)}$$

$$3. A \rightarrow B: \{m\}_{K(B)}$$

In the first message, A sends an encrypted message under its private key to B. Next, B encrypts this message under its secret key and sends it back to A. Finally, A decrypts the received message and sends the decryption result to B. When the third message is received, B decrypts it and obtains the secret message.
In this protocol the encryption and decryption operations are the same, and they are achieved by Xoring the message with the key.

3.1.2 ADVANCED ENCRYPTION STANDARD

The iterated symmetric cipher known as the Advanced Encryption Standard (AES) was created by Joan Daemen and Vincent Rijmen (Rijndael) [21]. The length of the cryptographic key for this cipher could be 128-bit, 192-bit, or 256-bit. Depending on the used key's size, the AES cipher encrypts and decrypts data in blocks of 128 bits, 192 bits, or 256 bits. To be more precise, AES uses 10, 12, and 14 rounds for 128-bit keys, 192-bit keys, and 256-bit key, respectively. Till now, no significant security attack against the AES has demonstrated successful.

3.2 K-NEAREST NEIGHBOR CLASSIFIER

One of the most straightforward machine learning algorithms is K-Nearest Neighbor (KNN), which is relied on supervised learning. The KNN algorithm inserts a new item in the classifier that most resembles the existing classifiers, presuming that the new data item and the existing ones are similar. The KNN approach preserves all the available data items and classifies a new data item based on how similar it is to the current classifiers. KNN is frequently used to solve classification issues [9].It can be conducted using the following steps:

(1) Identify the number of K neighbors.
(2) For each data item in the testing dataset

(2.1) Compute the Euclidean distance between testing item and all training dataset items.
(2.2) Sort the distances and assign them to an ordered list.
(2.3) Pick the first K items in the sorted list.
(2.4) Label the tested item according to the majority of classes that are present in the picked items.

3.3 RANDOM FOREST

A popular algorithm for classifying and predicting data is called "Random Forest", which is a type of supervised learning [10][22]. To increase the predictive accuracy of a dataset, this classifier constructs many decision trees on different subsets of that dataset. Expressly, Random Forest does not solely rely on a single decision tree, but rather averages the predictions from each tree or takes the majority of votes related to the predictions. Higher accuracy is achieved and the problem of over fitting is avoided by increasing the number of trees in the forest. In addition, the algorithm is trustworthy. Typically, a new data point may only have a minor effect on one tree, therefore the algorithm as a whole is not significantly affected [23]. This algorithm is fulfilled through the below steps:

(1) Choose random samples of data from a given training dataset.
(2) Create a decision tree for every training data sample.
(3) Determine the results of vote by averaging or maximizing the decision trees.
(4) Choose the predicted result that received the most votes as the chosen outcome.

3.4 PERFORMANCE METRICS

In this work, four assessment metrics are used: accuracy, precision, sensitivity, and F1_score [24][25].The accuracy metric determines how many times a classifier produced a correct classification results across the whole dataset. It is expressed as the ratio between the number of correct classified cases to the total number of the classified cases. While the ratio between the number of only correct classified positive cases to all of the cases that were either correctly or incorrectly classified as positive is denoted as precision. In other words, the times of the positive cases that are flawlessly classified are measured by the precision. Sensitivity measures how well the system can classify positive cases by dividing the number of rightly classified positive cases by the total number of positive cases together with the faulty negative cases. The precision and sensitivity metrics are combined to create the F1_score metric in order to assess a classifier's correctness. The above four metrics are provided in the equations below:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (2)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (3)$$

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (4)$$

$$F1\_score = \frac{2*Precision*Sensitivity}{Precision+Sensitivity} \quad (5)$$

4. PROPOSED SYSTEM

This paper proposes a secured system for heart disease prediction that serves the medical sector via presenting a pair of services: safeguarding the patient's formal and medical data and ensuring accurate diagnosis and recognition of disease. In essence, our system is composed of two phases, which are shown in Fig. 1 security provision and patient condition diagnosis. These phases will be illustrated below in depth.

4.1 SECURITY PROVISION PHASE

As indicated previously this phase aims at keeping the privacy of data related to the patient's medical file. To achieve that, we employ the AES encryption method to securely encrypt the patient's data. However, the main issue associated with this method is distributing the encryption key, i.e., the key must be securely exchanged between the involved participants.

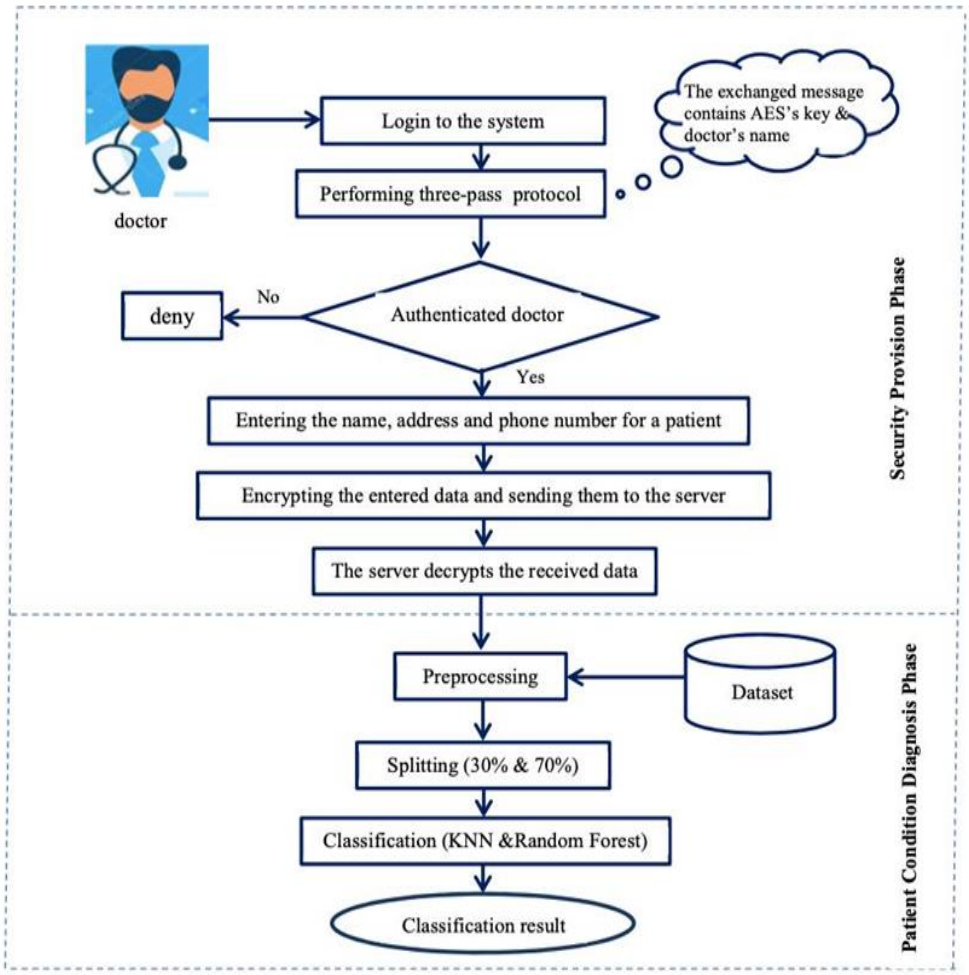

FIGURE 1. Architecture of the Proposed System

Under TP protocol, any user can exchange the encryption key without sharing any private information. At the same time, this protocol enables the system to inspect whether a user is an authenticated one by asking him/her to exchange the key and his/her name. Typically, at the end of the protocol, the server will check the received message which contains: the AES key and user's name, if they are identical to the saved ones, then the user is considered as an honest user, otherwise it is not. More precisely, the specification of TP protocol for this phase will be as follows:

$$1.\ A \rightarrow B: \{AES\ key, User's\ name\}_{K(A)}$$

$$2.\ B \rightarrow A: \{AES\ key, User's\ name\}_{K(A)}\}_{K(B)}$$

$$3.\ A \rightarrow B: \{AES\ key, User's\ name\}_{K(B)}$$

After a proper authentication, the system allow the user to inquire about an patient through sending to the server an encrypted message with AES key containing the patient's name, address, and phone number. Once the message is received, the server will decrypt it with the distributed key, and that will in turn attain the second authentication action. Moving to the next phase is only possible when the decryption process succeeds.

4.2 PATIENT CONDITION DIAGNOSIS PHASE

The primary aim of this phase is to choose an appropriate algorithm for accurately predicting whether a patient has heart disease or not. It relies on two popular machine-learning algorithms, namely, KNN and Random Forest, and after reformulating them to be suitable for the heart disease prediction issue. These algorithms exercise a set of clinical signs, including age, gender, chest pain type, blood pressure, serum cholesterol, fasting blood sugar, electrocardiography, maximum heart rate, exercise induced angina, ST depression, and the number of major vessels. The patient condition prediction phase is composed of the following stages:

4.2.1 PRE-PROCESSING STAGE

As our system uses an accessible dataset, there is a possibility that the data values in this set are manually entered, or collected from multiple sources and released by various official bodies. As a result, these values need to be preprocessed before being used for classification. Although there are different methods for performing the pre-processing action, in our proposed system we use the Feature Scaling method. It is a method of distributing the independent features present in the data uniformly within a constant range. It manages extremely variable magnitudes or values. In the absence of feature scaling, a machine learning algorithm would often rely on values regardless the units, so big values will be prioritized over small values even if these were actually bigger. A feature value is rescaled, using equation (6), as a part of the Standardization process, which yields a distribution with a mean equals to 0 value and a variance of 1.

$$V_{new} = \frac{V_i - V_\mu}{V_\sigma} \qquad (6)$$

Where V_i is a feature value in the dataset, V_{new} is a rescaled feature value, V_μ is the mean of the feature values, and V_σ is the standard deviation of the feature values.

4.2.2 SPLITTING DATA STAGE

Following sufficient preprocessing, the data will split into two sets, the training set, whose main objective is to enable the machine learning techniques to produce precise results, and the testing set that serves in assessing the system's performance. Essentially, 30% of the data is a test set, while 70% is a training set.

4.2.3 CLASSIFICATION STAGE

Due to the fact that the suggested system can directly affect human life and the diagnosis of health status, two popular machine learning classifiers: the KNN and the Random Forest, are compared in order to determine which one is the best depending on the evaluation metrics. Furthermore, the original algorithms of these classifiers are reformulated, such that they become more convenient to handle heart diseases matter. Algorithm 1 explains the steps of RF classifier, while those of KNN classifier is given in Algorithm 2.

Algorithm 1: The Reformulated Random Forest Classifier

Input: V_{ij}, where $i = 1,2,\dots,n$, and $j = 1,\dots,m$, which represents the feature values, n is the number of features, m is the number of values related to each feature, no_{trees} is the assumed number of trees, Fl is the vector of feature labels.

Output: The classified feature values $ClTV_{lj}$ intoCl_l class labels, such that $l = 1,2$, and Cl_1 is an intact person, while Cl_2 is a person with a heart disease.

Begin:

Step 1: Consider RV_{ij} as training set samples, which are randomly selected from the given input, TV_{ij} as testing set samples, which are also arbitrary selected.

Step 2: Classify the training set RV_{ij} into Cl_l, and save the results in $ClRV_{lij}$.

Step 3: Select TV_j from $TV_{ij}$to be a value that should be classified.

Step 4: Build a forest

For $p = 1\ to\ no_{trees}$

Identify the number of the subset features, such that $no_{sub-features} < n$, $no_{sub-features} = \sqrt{n}$

Randomly split the feature labels Fl into $no_{sub-features}$ subsets, and assign them to $SFl_{no_{trees},no_{sub-features}}$

For $q = 1\ to\ no_{sub-features}$

Arbitrary select values for $SFl_{p,q}$from $RV_{p,q}$ to be a random dataset $Y_{p,q}$

Construct a tree using $SFl_{p,q}$ and $Y_{p,q}$

End For

Use $SFl_{p,q}$ and $Y_{p,q}$to produce a decision **(**Cl_l) related to each constructed tree

End For

Step 5: Predict the class of TV_j

For each column in TV_{ij}

For $p = 1\ to\ no_{trees}$

If a column match the constructed tree **Then**

Calculate the majority of decisions related to the constructed trees

End if

End For

$ClTV_{lj}$ = the calculated majority of decisions

End For

Algorithm 2: The Reformulated K-Nearest Neighbor Classifier

Input: V_{ij}, where $i = 1,2,\dots,n$, and $j = 1,\dots,m$, which represents the feature values, n is the number of features, and m is the number of values related to each feature.

Output: The classified feature values $ClTV_{lj}$ intoCl_l class labels, such that $l = 1,2$, and Cl_1 is an intact person, while Cl_2 is a person with a heart disease.

Begin:

Step 1: Consider RV_{ij} as training set samples, which are randomly selected from the given input, TV_{ij} as testing set samples, which are also arbitrary selected.

Step 2: Classify the training set RV_{ij} into Cl_l, and save the results in $ClRV_{lij}$.

Step 3: Select TV_j from $TV_{ij}$to be a value that should be classified.

Step 4: Compute the Euclidean distance between TV_j and RV_{ij} using the following equation:

For $i = 1\ to\ n$

For $j = 1\ to\ m$

$$d(TV_j, RV_{ij}) = \sqrt{\sum_{j=1}^{m} (RV_{ij} - TV_{ij})^2}$$

Step 5: In a non-decreasing order, sort the computed Euclidean distances.

Step 6: Choose the number of neighbors and save it in k.

Step 7: From the sorted distances, select k values and put them in $d(TV_j, RV_{ij})_k$.

Step 8: Predict the class of TV_j

$no_{Cl_1} = 0, no_{Cl_2} = 0$

For $p = 1\ to\ k$

If RV_{ij} (which $\in$ $d(TV_j, RV_{ij})_k$) has $ClRV_{1ij}$**Then**

$no_{Cl_1} = no_{Cl_1} + 1$

Else

$no_{Cl_2} = no_{Cl_2} + 1$

End if

End For

If $no_{Cl_1} > no_{Cl_2}$ **Then**

Put TV_j in $ClTV_{1j}$

Else

Put TV_j in $ClTV_{2j}$

End if

5. EXPERIMENTAL RESULTS AND DISCUSSION

As our proposed system consists of two phases: security provision and patient condition prediction, this section focuses on analyzing and discussing the results of these phases. Note that, the system is implemented using Python language. The dataset of the first phase is collected manually from random patients at the Murjan Hospital in Hilla. Essentially, this database contains personal information, such as: patient's name, phone number, and address, see table 1.

Table 1. The database using in encryption layer

Feature	Explanation	Measurement
Name	patient name	Alphabet (A…Z)
Address	The patient's residential address	Name of city
Phone Number	The patient's mobile number	Number

Whereas, the second phase utilizes an accessible dataset[1] that consists of many characteristics, like age, gender, anemia, diabetes, high blood pressure, etc., which can significantly affect the occurrence of heart disease in individuals.

([1]) The dataset is available online at (https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci)

The first phase highlights the main problem of AES algorithm, which is the key distribution, and suggests the solution via using a Three-Pass protocol. In addition, using this protocol and AES encryption helps to achieve dual authentication for any user trying to access the system. However, security sides of neither the AES, nor the protocol had been demonstrated. The main reasons for that are: the AES never have been compromised before, and the Three Pass Protocol is susceptible only to man-in-the-middle attack where the attacker figures out that the transmitted data is encrypted using XOR logic.

Table 2. The performance results of Random Forest classification technique

The used technique and its setting	Number of trees	Accuracy	Precision	Sensitivity	F1_score
Random Forest with class **1** & no. of trees = 2 , 4 , 6, 8	2	0.93	0.87	1	0.93
	3	0.95	0.95	0.92	0.94
	4	0.93	0.87	1	0.93
	6	0.99	0.97	1	0.99
Random Forest with class **2** & no. of trees = 2 , 4 , 6, 8	2	0.93	1	0.88	0.94
	3	0.95	0.94	0.96	0.95
	4	0.93	1	0.88	0.94
	6	0.99	1	0.98	0.99

In the second phase, K-Nearest Neighbor and Random Forest classifiers are employed to categorize individuals as having heart disease or not. Each classifier is applied individually to the same data set of patients. To assess the performance of the classifier, a set of measurement metrics, such as accuracy, precision, sensitivity, and F1 score [20,21], is exercised. The obtained results indicate that the Random Forest method is superior to the K-Nearest Neighbor algorithm in terms of the used metrics. Table 2 shows the performance results of Random Forest classification technique, when the number of trees = 2 ,3, 4, 6. From this table, we can see that the best value of these metrics are achieved when the number of trees is 6.

Fig. 2 shows a set of experiments that are carried out on the RF algorithm and indicated their performance by changing the value of n_estimators (number of trees), such that the best results is attained when (n_estimators) = 6. While the results of K-Nearest Neighbor classifier are shown in Table 2, such that the number of neighbors (K) is chosen equal to 3, 5 , 7 , 9. It was found that the best result for the selected metrics yields when the value of K is 5. Fig. 3 shows a set of experiments that are conducted to state how the performance of KNN algorithm can be affected by different values of K. Typically, the best results are obtained when K=5.

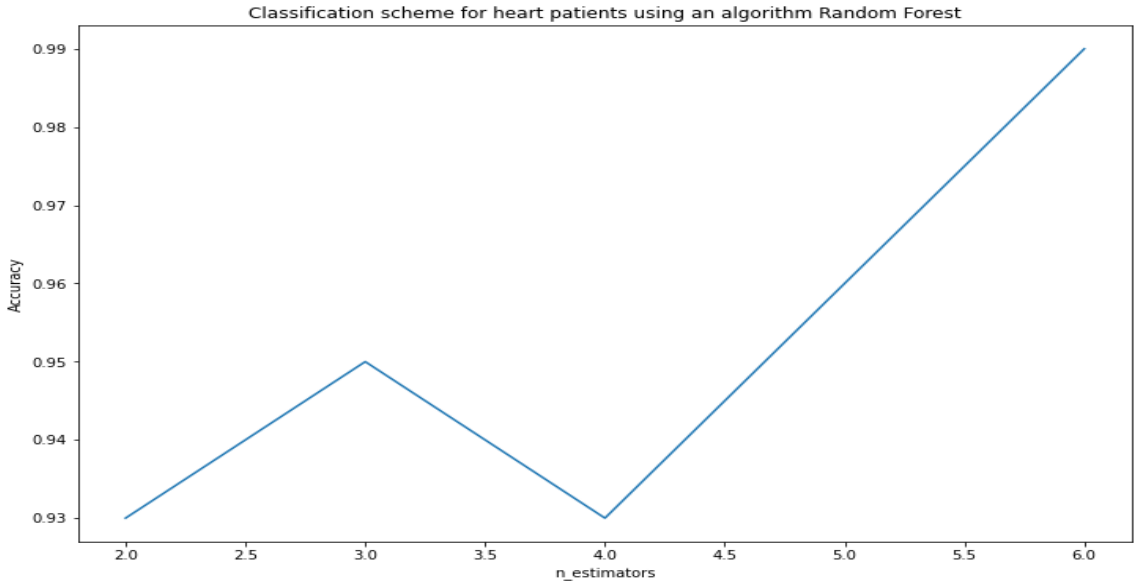

FIGURE 2. The accuracy of Random Forest with respect to the number of trees

Table 3. The performance results of K-Nearest Neighbor classification technique

The used technique and its setting	Number of trees	Accuracy	Precision	Sensitivity	F1_score
K-Nearest Neighbor with class **1** & no. of neighbors = 3 , 5 ,7,9	3	0.95	0.95	0.92	0.94
	5	0.97	0.97	0.95	0.96
	7	0.96	0.95	0.95	0.95
	9	0.96	0.95	0.95	0.95
K-Nearest Neighbor with class **2**& no. of neighbors = 3 , 5 ,7,9	3	0.95	0.94	0.96	0.95
	5	0.97	0.96	0.98	0.97
	7	0.96	0.96	0.96	0.96
	9	0.96	0.96	0.96	0.96

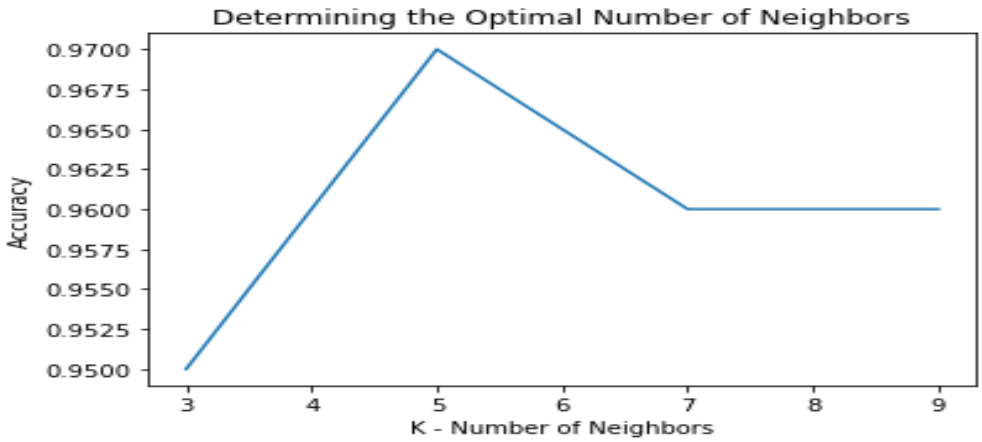

FIGURE 3. The accuracy of K-Nearest Neighbor with respect to the number of neighbors

Fig. 4 depicts a comparison between K-Nearest Neighbor algorithm and Random Forest algorithm depending on the adopted metrics set in this work. This figure shows the superiority of the Random Forest algorithm over the KNN algorithm.

Table 4 represents a comparison between the results of our work and those of some related works. This table consists of five columns: the reference number, the year of publishing this reference, the dataset that have been used by this reference, the technique that has been applied, and the achieved accuracy.

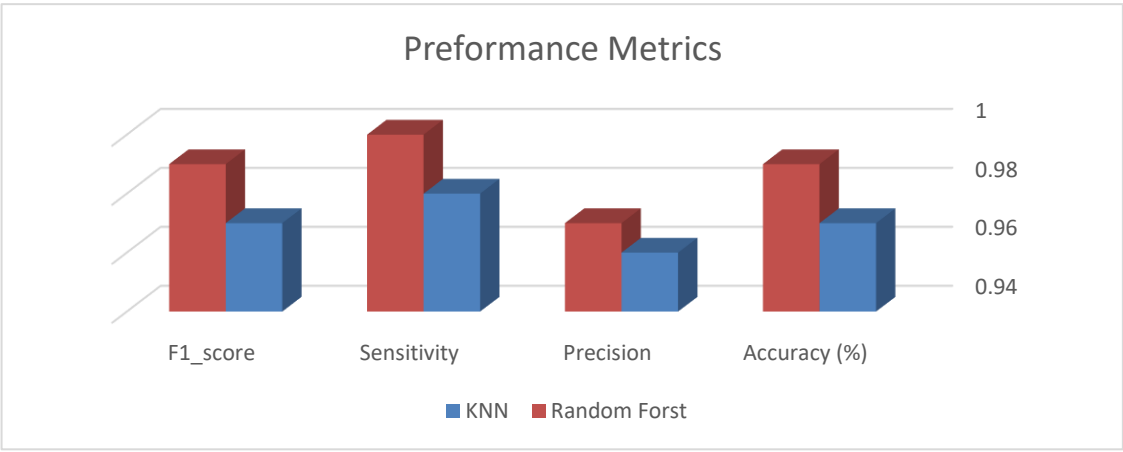

FIGURE 4. The performance comparison between RF & KNN

Table 4. Comparison of the proposed method with the results of some related work

Reference	Year	Dataset	Technique(s)	Accuracy
[11]	2019	medical profiles	Naive Bayesian (NB) classification and AES	89.77%
[19]	2021	University of California's open repository	RSD-ANN	96.30%
[17]	2022	UCI	XGBoost	92.19%
[18]	2022	UCI	Random Forest	96.28%
Our Proposed work	2023	www.kaggle.com	AES encryption, KNN , Random Forest	KNN = 97% RF = 99%

6. CONCLUSION

As a result of the Three Pass Protocol's use and the fact that a database was created for authorized users and their movements were tracked by the system, data is preserved when an authorized person is used to operate on the health system. Giving power to work by putting up two algorithms and selecting the best one, as the Forest Tree algorithm was accepted and produced outcomes that were superior to those of the KNN algorithm. To improve performance and select the optimal parameter, however, more than one parameter was examined in the two methods. Future research should consider the potential effects of launching attacks against the Three Pass Protocol more carefully.

Funding

None

ACKNOLEDGEMENT

The author we would like to thank the reviewers for their valuable contribution in the publication of this paper.

CONFLICTS OF INTEREST

The author declares no conflict of interest

REFRENCES

[1] F. Z. Patrick Rossignol, Adrian F. Hernandez, Scott D. Solomon, "Heart Failure Drug Treatment."
[2] N. Dutta, U. Subramaniam, and S. Padmanaban, "International Meeting on Advanced Technologies in Energy Mathematical models of classification algorithm of Machine learning." pp. 2018–2019, 2019.
[3] V. V Aishwarya Mujumdar, "Diabetes Prediction using Machine Learning Algorithms."
[4] L. dwight Smith, "Cryptography; The Science of Secret Writing," *Think of a Number*. pp. 71–80, 2021. doi: 10.1201/9781003072362-8.
[5] A. J. M. P. C. van O. S. A. Vanstone, "HandbookOfAppliedCryptography_AMenezes"
[6] A. A. H. and A. A. M. M. Haque, "A_Comparative_Study_of_the_Performance_a.pdf."
[7] Khasanah, P. T. Nguyen, G. Gunawan, and R. Rahim, "Three-pass protocol scheme on vigenere cipher to avoid key Distribution," *Journal of Critical Reviews*, vol. 7, no. 1. pp. 68–71, 2020. doi: 10.22159/jcr.07.01.13.
[8] R. L. R. and L. M. A. Adi ShalJlir, "Mental Poker"
[9] A. P. E. H. T. M. COVER, "Nearest_neighbor_pattern_classification"
[10] I. Barandiaran, "The Random Subspace Method for Constructing Decision Forests."
[11] A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, "Design and implementing heart disease prediction using naives Bayesian," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, vol. 2019-April, no. Icoei, pp. 292–297, 2019, doi: 10.1109/icoei.2019.8862604.
[12] Mukhayo Rasulova and Jakhongir Yunusov Institute, "Definition of a Three-Pass Protocol using the Lieb-Liniger Model."
[13] D. Vashi, H. B. Bhadka, K. Patel, and S. Garg, "An Efficient Hybrid Approach of Attribute Based Encryption for Privacy Preserving Through Horizontally Partitioned Data," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 2437–2444, 2020, doi: 10.1016/j.procs.2020.03.296.
[14] D. S. and P. C. Umarani Nagavelli, "Machine Learning Technology-Based Heart Disease Detection Models."

[15] X. Yang, R. Lu, J. Shao, X. Tang, and H. Yang, "An efficient and privacy-preserving disease risk prediction scheme for E-healthcare," *IEEE Internet of Things Journal*, vol. 6, no. 2. pp. 3284–3297, 2019. doi: 10.1109/JIOT.2018.2882224.

[16] P. G. A. Ankita Jagtap, Arbaaz Bebal, Nomit Bhatnagar, Pratiksha Kamthe, "Blockchain-based Secure Healthcare for Cardio Disease Prediction."

[17] W. B. G. & M. Z. D. Gizeaddis Lamesgin Simegn and To, "Computer Aided Decision Support System for Diagnosis of Heart Diseases.pdf."

[18] S. S. U. Ch. Anwar ul Hassan, Jawaid Iqbal, Rizwana Irfan, Saddam Hussain, Abeer D. Algarni, Syed Sabir Hussain Bukhari, Nazik Alturki, "Effectively Predicting the Presence of Coronary Heart Disease"

[19] M. B. Abdulaziz Albahr , Marwan Albahar , Mohammed Thanoon, "Computational Learning Model for Prediction of Heart Disease"

[20] J. C. and J. Jacob, "A Survey of Authentication Mechanisms," no. c. pp. 1–60, 2010.

[21] J. Daemen and V. Rijmen, "AES proposal: Rijndael." 1999.

[22] Saba Abdulbaqi Salman, Saad Ahmed Dheyab, Qusay Medhat Salih, & Waleed A. Hammood. (2023). Parallel Machine Learning Algorithms. Mesopotamian Journal of Big Data, 2023, 13–17. https://doi.org/10.58496/MJBD/2023/002

[23] K. J. G. S. Dmytro Chumachenko , Mykola Butkevych , Daniel Lode , Marcus Frohme and and Alina Nechyporenko, "machine learning methods in Predicting Patients with Suspected Myocardial Infraction Based on Short-Time HRV Data"

[24] H. K. Obayes, F. S. Al-Turaihi, and K. H. Alhussayni, "Sentiment classification of user's reviews on drugs based on global vectors for word representation and bidirectional long short-term memory recurrent neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1. pp. 345–353, 2021. doi: 10.11591/ijeecs.v23.i1.pp345-353.

[25] I. B. D. Lotfi Mhamdi, Oussama Dammak, François Cottin, "Artificial Intelligence for Cardiac Diseases Diagnosis and Prediction Using ECG Images on Embedded Systems."