International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15 (2023) pp. 616-627

Securing Medical Data with Modified AES and Machine Learning Classifiers Based on Meta-heuristic Feature Selection Algorithms

[*1]Uma K, [2]Hanumanthappa M

[*1]Research Scholar, Department of Computer Science and Applications,
Bangalore University, Bangalore, India
Veereshsajjan.uma@gmail.com

[2]Professor, Department of Computer Science and Applications,
Bangalore University, Bangalore, India

Abstract: **Security is a major concern in all automated applications and particularly in medical field, secure data transmission is significant in order to preserve the health of patients. To provide integrity to medical data, numerous cryptographic algorithms have been utilized for security purposes. But, it faced complications in security while handling larger volume of data and increased execution time taken for the process of both encryption and decryption which lowers the performance of system. In order to overcome such challenges, the proposed work is designed on two stage implementation. The first phase is providing security with modified AES (Advanced Encryption Standard) algorithm on medical dataset. The patient information are encrypted using AES involved with advanced bit permutation operation for improving the encryption efficiency. The extraction of data at the receiver end can be performed only by the authorized access. The second phase takes input as encrypted data and performs classification for predicting the disease with the utilization of three meta-heuristic algorithms such as GA (Genetic Algorithm), PSO (Particle Swarm Optimization) and GWO (Grey Wolf Optimization) separately and utilizes the XGBoost classifier to each algorithm individually for obtaining the optimal features from each of the algorithms. Finally, k-fold cross validation process predicts the efficiency level of classification in all algorithms. Based on data in Cleveland dataset, the disease of patients can be categorized with the present research. The experimental internal evaluations and the comparative analysis exhibits the efficiency of the system in terms of better security, minimum execution time and classification accuracy.**

Keywords: AES (Advanced Encryption Standard), GA (Genetic Algorithm), GWO (Grey Wolf Optimization), XGBoost classifier, PSO (Particle Swarm Optimization).

I. Introduction

The transmission of health data is included with clinical analysis results and reports which is required in hospitals for reducing the repetition of test to be taken for patients. The faster sharing of clinical data increases the treatment for patients [1]. Hence, computerized systems are utilized for exchanging health information in a faster way. Subsequently, securing those beneficial data is equally significant. Numerous studies have been made on securing the transmission of health oriented information with cryptographic methodologies. The encryption and decryption mechanism has provided security keys for data monitoring and retrieval process by authorized persons[2]. Identification of appropriate keys have been found to be a difficult task but, then hackers tried to extract confidential data and misused the medical information. The considered method [3] has provided security to data by means of bit-mask based genetic algorithm. It has considerably reduced the replication of medical data which has been transferred across various medical organizations. The optimization algorithm prevents the method to get locked at local optimum. Boolean technique used in the recommended method has avoided premature convergence with the integration of cryptographic features. The encrypted data has been extracted at the receiver end through reverse process of the cryptographic functions. The potentiality of the algorithm has been proved through performance analysis in terms of PSNR (Peak Signal to Noise Ratio) and MSE (Mean Square Error). Even though several studies have shown efficiency in data protection techniques, secure communication between devices have been emerging as a barrier for healthcare systems due to the purpose of managing diagnostic data about patients [4]. Threats by intruders affect the reliability, accuracy and integrity of the system. The considered approach [5] has utilized an encryption process that hides the private data with the utilization of shared key and scrambling matrix which has created cryptography in two ECG signals. In order to extract the secret information of affected person from two ECG signals, the method of decryption has used the same share key [6]. The authorized clinicians holding the key have access to the data and has the potential in reconstruction of the original data. For providing protection to larger data, the security mechanism needs the scalability and has suffered with computational overhead. The recommended research [7] has taken efforts on storing large volume of medical data in cloud and has provided

protection mechanisms in cloud with less computational time. In order to achieve monitoring and evaluation operations effectively, aggregation of medical data is considered to be important [8]. The suggested methodology [9] has utilized the block chain based technique in both the user and the clinicians systems. AES based cryptographic function has performed block chaining and cryptic functioning with the utilization of hash keys. It also has been included with the DL (Deep Learning) algorithm for diagnosing the disease of patients. AES- 128 and SHA-256 have been utilized for cryptic functioning over the data. It has used RNN technique for diagnosing and prediction of particular disease with the information in the dataset. The security parameters such as the computational time and processing capacity with blocks have proved the effectiveness of the system. Even though traditional studies have provided many techniques and methodologies for effective protection of data in healthcare systems, time consumption in encryption and decryption approach needs improvement. Moreover, security framework with efficient data retrieval alone has not helped the patients in recovery. To this purpose, certain efficient classification algorithm with greater level of accuracy is required. The main intention of the proposed research is to address the challenges of the existing studies in security. Based on the analysis made from various approaches, the aim and objective of the study is listed as follows:

- To achieve security to medical data with modified AES mechanism involoved with bit permutation function for encrypting the entire information in dataset.
- To accomplish the task of classification with the GA (Genetic Algorithm), PSO (Particle Swarm Optimization) and GWO (Grey Wolf Optimization) algorithm for diagnosing the disease with the encrypted data.
- To perform selection of optimal features with the integration of XGBoost algorithm with the three metaheuristic algorithms indivdually.
- To evaluate the system with performance metrics for predicting the efficient meta-heuristic algorithm with graeater level of accuracy.

A. Paper Organization

The review paper is organized as follows. Section 2 shows the review made on several existing studies related to security. Section 3 outlines on the proposed security mechanisms and the classification strategies utilized in the study. Section 4 provides the result analysis based on the performance of the system evaluated with metrics and comparative analysis proves the effectiveness of the proposed model compared to the existing security frameworks. Finally, the proposed work is concluded at Section 5.

II. Literature review

Internet applications have reduced the involvement of humans in performing numerous tasks. Such internet based devices in medical field requires large amount of data and transmission of medical data needs real-time computation and data storage [10]. People need security and privacy for storing and accessing the medical health records. The health related sensitive information stored in EHR (Electronic Health Records) is being hacked by unauthorized access and creates a threat to patients. The suggested approach [11] has concentrated on the encryption of the database before the process of outsourcing the data in cloud in order to increase efficiency and security performance of the data stored in medical database. The considered research has taken three encryption based methodologies like AES, DES and Diffie-Hellman approach for encrypting the patient information based on the size of the file. From the inferences obtained through evaluations performed on three techniques, AES has been considered to be efficient in terms of providing protection. The medical data transmitted through cloud and unreliable networks are vulnerable to threats and cryptographic techniques have provided sufficient techniques to deal with the issue [12]. The RSA (Rivest Shamir Adleman) based encryption method has been utilized in the suggested study [13] for transmitting safer data and the performance evaluation of the system has revealed that it has the potentiality in preventing RS attacks. The considered algorithm has provided confidentiality and prevents misuse of clinical data.

Technological advancements in medical field is increasing and changes in the clinical therapies urges the need for historical medical related information about patients [14]. The requirement of such information also needs for protecting the patient report from attackers. The health care information storage system in the recommended study [15] has protected clinical data through utilization of block chain technique. It has been implemented with three sections such as data authentication process in initial stage. It has been performed through quantum cryptography method [16]. Followed by that, encryption phase which has been handled with AES algorithm. The last stage of data retrieval method has been executed with the SHA cryptographic algorithm in order to provide resistance against frequent data attacks. It has ensured protection of patient data and has maintained security in health report systems effectively [17]. The three considered parameters utilized in the recommended technique were scalability, integrity and access control. The time taken for verification, recovering and including information into the database has been calculated for predicting the scalability nature of the system. The integrity of the system has been validated since no other individual has been allowed for altering the patient data. Only the authorized persons have been given complete authority [18].

Most of the medical centers store patient data for further medical processing procedures. It poses impacts on providing secure data. Missing of any medical information about a particular patient highly affects the patients' health

[19]. The suggested approach [20] has taken efforts in determining security in hospital database. It has utilized AES and made enhancement in the algorithm and has named the technique as P-AES. It has been integrated with the RSA and considerably increased the speed of encryption and decryption with better computation ability [21] and has solved the security issues to some extent and validated the efficiency of the system with the clinical information management system [22]. Many conventional data encryption algorithms have been focused from the view on security based architecture and has lacked the users comfort zone in terms of protection [23]. Since such types of security related algorithms have transferred the protection of data on to the security of keys, then confidentiality of data has been considered to be a threat when keys have been exposed. The considered technique in [24] has designed the security mechanism with the selective prediction of encryption method with the fragmentation process and followed by dispersion technique for providing security on data even when both the data transmission medium and keys have been compromised. It has been structured mainly based on users and the efficiency of the system has been validated. The selective method of encryption has protected data shared in EHR. The fragmentation technique utilized in the study has provided additional protection over various fragments of storage and the resisting approach has created trust among the users [25]. It has implementation on Health Industry 4.0 that solved threats even when both keys and data stored in EHR have been leaked. The effectiveness of the algorithm has been validated by determining the level of protection provided.

The data security in fog computing has provided consistency and security to a certain level based on the nodes structured at the edges and has reduced the rate of latency which has been considered to be a significant factor in the medical data protection [26]. The recommended study [27] has provided AES algorithm for encryption which has enhanced the security parameter in fog computing. The challenges to security in fog computing have been solved through reliable real time storage and retrieval process with effective protection using AES security algorithm. The possibilities of tampering of data, forgery, leakage and threat attacks while sharing data over public and cloud environments have been considered to be the major issue in information and network technology. Techniques like concealment, decentralization and immutability in block chain method has provided solutions to the security problems. The considered study [28] has provided complete medical information framework for safer storage and transmission of clinical data. The data collection system in the recommended study has collected data from various sources of real-time patient medical record needed for surgery. It has implemented concealed data sharing method based on proxy-re-encryption algorithm for enhancing security. The evaluations made on safety and access control of the system has provided ways for performing remote medical treatment.

A. Problem identification

The limitations with respect to security has been analyzed from the existing studies and are listed as follows

- The block chain security model in the health care system has provided access to all clinicians which leads leakage of data. It needs to exhibit access only to the trusted medical care facilities. The scalability problem addressed in the considered study can be tackled with the artificial intelligence algorithms for effective analysis of diseases from patient's data with minimal execution time [29].
- Protection mechanism with selective encryption technique with fragmentation on storage has concentrated on providing confidentiality to specific fragments. So, when DL algorithms have been applied on the database to extract information for diagnosis of diseases, it leads to data leakage and computational overhead persisted. It needs to be resolved by enhancing the encryption algorithm [30].

III. Proposed Methodology

Medical data stored in database are prone to vulnerability and cryptographic functions like AES is considered to be the frequently utilized encryption algorithm which provides security to data stored in datasets. However, it takes more time for encryption and suffered with computational overhead. Hence, the proposed system enhances the AES algorithm with the integration of bit permutation into the steps of AES in order to overcome the computational complexities and also for making the encryption process faster. The modified AES technique replaces the mixcolumn function in the AES technique with the bit permutation operation which does not involve complex computation and finds at ease for implementation. The cryptographic modified-AES algorithm in the proposed research is being utilized for providing security to electronic medical health data during transmission. It is involved with symmetric block cipher which has the potential to perform encryption with encipher and decryption with decipher. Encryption is the process of converting the original text data into unknown form of unpredictable text called as the cipher text and the decryption process converts the cipher text back to its real data. The enhanced form of AES is being designed to be operated on a larger amount of data with increasing execution time. The modification is being done with the bit permutation process for enlarging the performance of encryption. Such optimization in AES improves efficiency of encryption method which thereby provides security to data from intruders attack. The architecture diagram of the proposed model is given in Figure.1

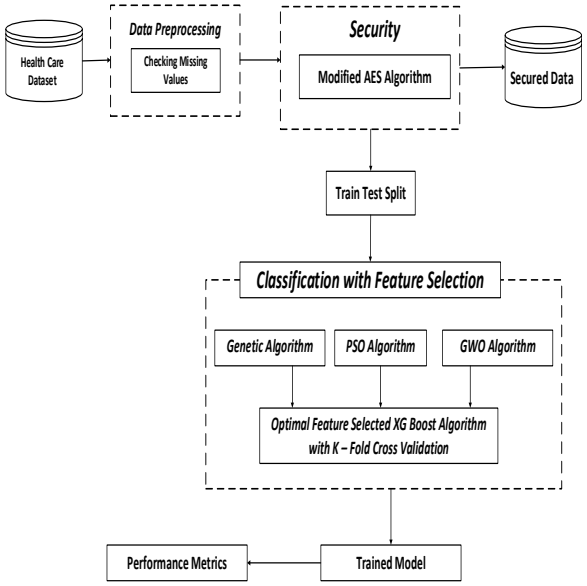


Figure.1 Overall Process flow of the Proposed System

The proposed security framework provides integrity and confidentiality for the digital data being transmitted. From Figure.1, it is clear about the step by step procedure of the proposed flow. The process initially starts with the data pre-processing stage. It takes input data from the heart healthcare dataset. Missing values and improper, unrelated data from the dataset is being removed and clear, finite amount of data are retrieved after pre-processing. Such cleaned data is provided protection using modified AES algorithm. After, the data is being secured and encrypted, it moves on to the classification phase for diagnosing the disease from the information obtained from the dataset. The proposed system uses three meta-heuristic algorithms such as GA, PSO and GWO for performing feature selection process. After the features are selected individually from three algorithms, it undergoes classification process with XGBoost algorithm. All the algorithms are validated with k-fold cross validation in order to determine the accuracy level of each type of algorithm. The security efficiency is being evaluated with key length and compared with other algorithms for exhibiting the effectiveness of the proposed system. The algorithm with greater classification accuracy is potential in diagnosing and classifying the type of disease effectively.

A. Encryption and Decryption with Modified AES

Modified AES is being developed in the present study to provide security to text data stored in EHR of patients specifically. The main intention of this algorithm is to provide minimum computational time with better level of security. It takes the 128 bit key length considered as the array of bytes and forms a matrix called as states. The number of rounds in AES is based on the key length in which states are being subjected to transformations for encrypting the plaintext into cipher text. The general AES algorithm is involved with four kinds of transformations such as Add round key, Sub bytes, Shiftrows and Mixcolumns. Since the last function of AES namely Mix-column has greater computational load and expensive due to complex calculations, it requires additional resources for implementation. It also has made the execution slower. Hence, Mixcolumn is being replaced with the Bit permutation operation to make execution faster. 128 bit block is taken as input which takes 4 x 4 matrix of 16 bytes. The states are changing at every phase of encryption and decryption. The structural diagram of Modified-AES encryption algorithm is given in Figure.2

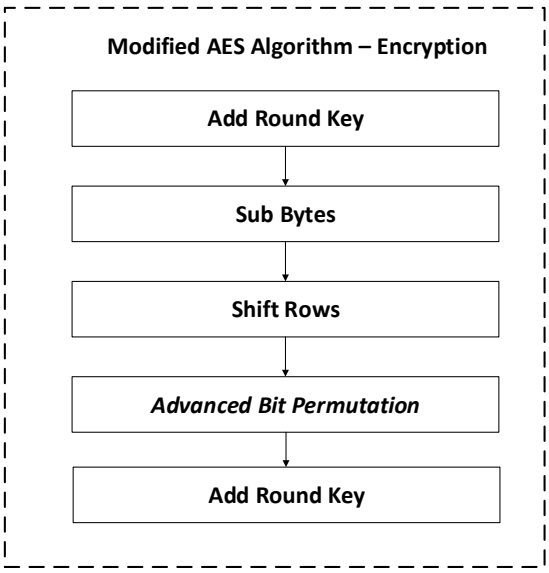


Figure.2 Block diagram of Modified AES

From Figure.2, it shows the modification of AES with the utilization of advanced Bit permutation operation. The importance of using this operation is it does not hold complex computations and involved only with the position shifting of bits at every state. The 128-bit key length taken in the proposed work is operated with 10 rounds of transformations. The stages in the modified security technique is involved with add round key, substitution bytes, shifting rows and advanced bit permutation is being utilized. The process of encryption is being applied to the database from first block and generates a key. With that key, the encryption of next block in dataset takes place and the process continues till the end of the database. Each and every encryption process generates a different key which assists in providing higher security level to the data from the attackers. The step by step progression of encryption with modified AES is given in pseudo code 1

Pseudo-code 1: Encryption algorithm with Modified AES
Input: Heart data from dataset Output: Encrypted dataset Initialize Begin Input: : plain text function AES (in = byte [128], out = byte[128], keyarrayroundkey[Nr + 1] byte state[128] state = 1

Add_Round key(state, roundkey[0])
For i = 1 to Nr − 1 do
SubBytes(state)
ShiftRows(state)
Add_Round key(state, roundkey[Nr])
Out = state
Return out Encrypted text
// Block to blockwise Encrypted Pseudo code//
Ciphertext
− →AES Encrypt (pwd, Heart data1, … … Heart datan .
AES Encrypt (pwd, heartdata)
− → Converted into block to block verification

From pseudo code 1, it is clear that the algorithm starts with the initial heart input 128 bit data and follows the four kinds of operation in modified AES in order to encrypt the text. Followed by that, block level encryption is being performed which encrypts the entire block with the key and using that, the next succeeding blocks in the database are also being encrypted. The column level encryption with the bit permutation operation effectively secures data in dataset. After the process of encrypting the text, the decryption needs to be taken place for extracting the original text data from the encrypted data and hence the work flow for decryption method is given in pseudo code 2.

Pseudo code 2: Decryption using modified AES

key − −→ $\text{Encrypt}_{\text{pwd}}$, time−→ Encrypt_password
for as much as tolerance give the secured pwd
if key = get_ secured pwd
key → Encrypt_pwd
plaintext → AES decrypt (Cipertext, key)
end if
end for
output plain text

The proposed work allows only the authorized users for extracting the real data. The pseudo code 2 exhibits the parameters used in the decryption process. Encrypted password key is to be given and if it accurately matches with the key being generated, then the process of decryption will be taken place with the function AES decrypt (Cipher-text, key).

B. Classification with Feature Selection

The encrypted data acquired from the modified security model is being induced into the classification framework as input data. It performs the classification process by diagnosing the disease based on certain criteria and predicts the output from the analysis on the medical information stored in the database. The clinicians utilize such records to proceed with further treatment process. If such health related important results have been missed from the database, it is likely to be difficult for predicting the type of disease and the necessary diagnosis needs to be taken again for proper treatment procedures. Hence, modified AES based encryption technique is followed for encrypting the medical data to provide protection from attackers and then classification is being performed for predicting the kind of disease. In classification approach, the proposed work utilizes three meta-heuristic algorithms such as GA, PSO and GWO separately for selecting features. After feature selection process, classification is being performed with the XGBoost classifiers that predicts the disease efficiently. Finally, k-fold cross validation operation is being utilized which brings out the better algorithm that suits for diagnosis. Initially, the optimal subset of instances are being selected through GA. It utilizes the population based searching technique for finding out the best solution for the problem and its operation is depicted in pseudo code 3

Pseudo code 3: Genetic Algorithm

Input : D is dataframe, D_i is the subset of N, POP
− is population, S_p
− selection probability
C_p crossover probabilty, M_p mutation probabilty. D-N, N-H
Output : selected features $S_i \subseteq N_i$
Randomly initialize a population of size H.
repeat
for $(i = 1; i \leq H; i = i + 1)$ **do**
calculate the fitness of $\mathbf{G_i}$
end
for $(i = 1; i \leq H; i = i + 1)$ **do**
Generate a random number $r \in [0,1]$
if$(r \leq g)$ **then**
Select G_i
end
else
if $(g_{i-1} \leq r \leq g_i)$**then**
select (G_i)
end
end
end
select individuals by c_P, and conduct crossover
select individuals by M_p, and conduct mutation
until (the termination condition is met):
output S_i

The fitness function calculated in the pseudo code 3 finds out the best offspring from parent and generates the next level of generation until termination. Finally, the best and optimal offspring being generated are considered to be the best solution obtained from the algorithm. The next classification is processed with the PSO. It is used for solving optimization problem and explores the solution space. It is computed with the swarm of elements called as particles and every particle identifies the candidate solution along with its co-ordinates in the D-dimensional space. The operation of PSO is given in the following pseudo code 4.

Pseudo code 4 : Particle Swarm Optimization

Input :: input features from dataset D
Output :: selected features from the pso algorithm
begin
params init
pop init
fitness function
$f(a) = a1^2 + a2^2 + a3^2$
while $(\text{no}_{\text{generation}}, \text{or criterion}_{\text{stop}}$ is not reached):

for $(i = 1$ to no of $A_{particles})$
{
if the $fitness_{values}$ of $A_i^t > P_{best(i)}$
then
update $P_{best} = A_i^t$
if the $fitness_{values}$ of $A_i^t > G_{best}$
then
update $G_{best} = A_i^t$
Update velocity particles $= Vel_i^t$
Update position particles $= vel_{i\,dim}^{t+1}$
Next particle
}
Next generation
}
return the selected set of features in the dataset

From the pseudo code 4, it is clear that, the position of i^{th} particles are represented by $A_1, A_2, \ldots A_n$ and the velocity of the particle is being denoted as V_{el} The fitness function is calculated for every particle in swarm and is being compared with the fitness of best result obtained from the previous outcome and the fitness of best-particles from all particles in swarm. The position and velocity of the particle is updated after predicting the best outcome by the following equation 1,

$$vel_{i\,dim}^{t+1} = w \times vel_{i\,dim}^{t} + ac_1 \times r \times (P_{idim} - a_{i\,dim}^{t}) + ac_1 \times r \times (P_{global\,dim} - a_{i\,dim}^{t}) \quad (1)$$

$$a_{i\,dim}^{t+1} = a_{i\,dim}^{t} + vel_{i\,dim}^{t+1} \quad (2)$$

Where dim=1,2,3,…..n and dim represents the dimensions of search space. w is denoted as the inertia weight which is being predefined and ac_1 is considered to be the acceleration constants. The parameter r is the random number uniformly generated within the interval. Through computation of swarm algorithm, best particle is being predicted. The final classification technique used in the study is GWO and its operating flow is given in pseudo code 5

Pseudo code 5: Grey Wolf Optimization

Input :: input features from dataset D
Output :: selected features from the grey algorithm
begin
params init
pop init
$a_i = (i = 1,2,3 \ldots n)$
init z, A and C
fitness fn for all search agent
alpha, beta, delta search agents (sa)
alpha(a) = best sa
beta(a) = second best sa
delta(a) = third best
while$(t < max_{iter})$
for sa
update the position of the current sa
end for
Update z, A and C
update alpha(a), beta(a) and delta(a)
$t = t + 1$
end while
return alpha(a) the best selected subset from datafrar

From the pseudo code 5, it is clear that, GWO uses the three types of wolves such as alpha, beta and delta for simulating the operation. The algorithm imitates the hunting nature of the wolves and figures out the best prey. Each kind of wolf finds for the location of prey and the three best location is being updated with the current best location. The optimal location of prey is being finalized with the alpha wolf through which best solution can be obtained. It is being employed for selecting best features from the medical dataset for performing diagnosis. After all three algorithms have been successfully predicted the best features, XGBoost algorithm is utilized individually with each algorithm to obtain optimal features from the dataset. The step by step process of XGBoost with k-fold cross validation is given in pseudo code 6.

Pseudo code 6: XG Boost Algorithm

Input data : dataset with k fold cross validation
initialize $f_0(x)$;
Step1: $feature_{selection}(data) = pso(f(s_1) + gene(f(s_2) + gw(f(s_3))$
Step 2: for $n = 1,2,3 \ldots \ldots, M$ do
Step3: calculate $i_n = feature_{selection}(data) + \frac{\partial L(g,f)}{\partial f}$;
Step 4: calculate $j_n = feature_{selection}(data) + \frac{\partial^2 L(g,f)}{\partial f^2}$;
Step 5: determine the structure by choosing the k fold
Step 6: for i in range $(dataset(df))$
Step 7: $rand_{part}$ data info:
Step 8: $train_{set}(i)$
Step 9: $test_{set}(i)$
Step10: $train_{set}(i) = k_{train} + k_{test}$

Step 11: $A = \frac{1}{2}\left[\frac{G_L^2}{H_L} + \frac{G_R^2}{H_R} - \frac{G^2}{H}\right] + train_{set}(i)$
Step 12: determine the leaf weights $mel^* = -\frac{G}{H}$
Step 13: Determine the base learner $\hat{b}\,(p) = \sum_{j=1}^{T} wI$
Step 14: add trees $f_n(x) = f_{n-1}(p) + \hat{b}\,(p)$;
Step 15: call train and test mean values
Step 16: end

From the pseudo code 6, the algorithm is projected from the additive models and frequently addressed with the base learners. The additive framework is the combination of base learners which is being computed through function given in equation 3,

$$f(a) = \sum_{n=1}^{M} f(b_n)(a) \quad (3)$$

Where f(a) is considered to be the sum of base learners and n=1,2,3,…M and M is considered to be the members of base learners. L is represented as the risk minimization parameter computed with $L = (f(a, b))$ and the equation for base learner is being computed with the following equation 4,

$$\text{argmin}_{base} \sum_D \left[f(b_n)(a)g(a,g) + \frac{1}{2}f(b_n)^2(a)h(a,g)\right] \quad (4)$$

where heart dataset $MD = \{(a, g)\}$is a dataset and the additive model is computed through equation 5,

$$I(a,g) = \frac{\partial L(f_{n-1}(a),g)}{\partial f} h(a,g) = \frac{\partial L(f_{n-1}(a),g)}{\partial f^2} \quad (5)$$

The updated outcome is being iterated with the boosting function and it is given in equation 6,

$$f_n(a) = f_{n-1}(a) + f(\widehat{b_n})\,(a) \quad (6)$$

The tree model of boosting algorithm is formulated in equation 7,

$$f(a) = \sum_{j=1}^{T} w_j I\,[a \in R_j] \quad (7)$$

where w_j is the constant fit in specific region R_j and I is considered to be the group of indices with input p and j^{th} leaf for j= 1,2,3,…T. The optimal prediction of leaf weights are equivalent to the learning of leaf weights. It requires the split which maximizes the gain and is considered to be the better optimization of k-fold validation and manipulated in following equation 8,

$$mel^* = \text{argmin}_w \sum_D \sum_{j=1}^{T} \left[g(a,g)w_j + \frac{1}{2}h(a,g)w_j^2\right] \quad (8)$$

$$\text{Where argmin}_w \sum_{j=1}^{T} \left[I\, w_n + \frac{1}{2}Jw_j^2\right] \quad (9)$$

$$\text{were } I = \sum_D i(a,g) \qquad J = \sum_D j(a,g)$$

The optimal leaf-weights are determined by the following equation 10,

$$mel^* = -\frac{I}{J} \quad (10)$$

The left and right binary splits maximize the gain and represented in equation 11,

$$W = \frac{1}{2}\left[\frac{I_L^2}{J_L} + \frac{I_R^2}{J_R} - \frac{I^2}{J}\right] \quad (11)$$

Where L and R represents the branches of the tree and optimized solution is being yielded from the algorithm.

IV. Results and Discussion

The performance of the proposed model is being evaluated with the encryption efficiency of text data along with the classification accuracy obtained from the algorithm. The following section describes the computation result and comparative analysis performed with other algorithms to predict the effectiveness of the proposed security structure.

A. Dataset Description

The dataset used in the proposed work is Cleveland dataset taken from the repository and consisted with numerous individual data with different features.

Dataset Link - https://archive.ics.uci.edu/ml/datasets/heart+disease

B. Performance analysis

The computational result obtained from each algorithm is being illustrated in this section. The outcomes obtained from GA is given in following Table.1

Performance of Feature selected by Genetic algorithm		
Optimized XGBoost classifier	Maximum accuracy	90.1639
	Minimum accuracy	75.4098
	Overall accuracy	81.2158
	Std deviation	0.04914
	Max precision	87.8787

Table.1 Performance evaluation of GA

From Table.1, it is clear that the maximum accuracy obtained from the computation of GA is 90.1639 and maximum precision value obtained through this algorithm is 87.8787. The graphical representation of GA is given in below Figure.3

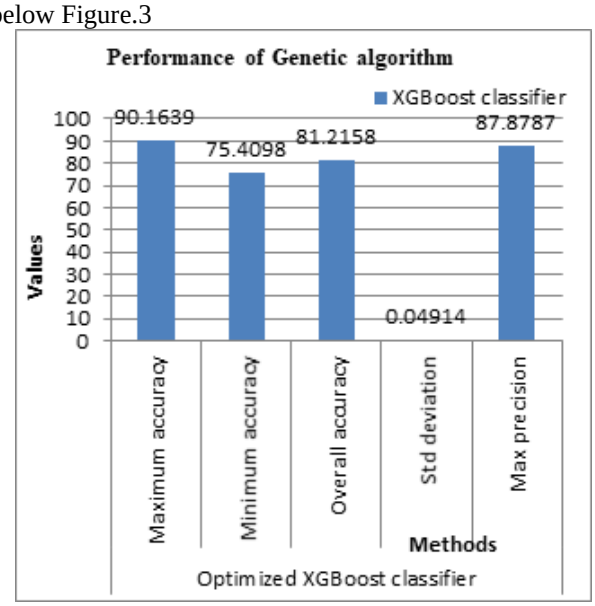


Figure.3 Performance metric validation of GA

From Figure.3, it is evident that considerable accuracy and precision value is being achieved through the implementation of GA in the process of feature selection and classification with XGBoost. The accuracy performance evaluated from GWO with XGBoost classifier is tabulated in the following Table. 2

Performance of Feature selected by Grey wolf algorithm		
Optimized XGBoost classifier	Maximum accuracy	85.2459
	Minimum accuracy	66.666
	Overall accuracy	77.9153
	Std deviation	0.06457
	Max precision	91.1764

Table.2 Classification with GWO

From Table.2, it is being inferred that level of accuracy obtained from the GWO is 85.2459 and the precision is calculated with 91.1764. It achieves better than GA in terms of precision and lower in accuracy and the graphical representation is given below in Figure.4

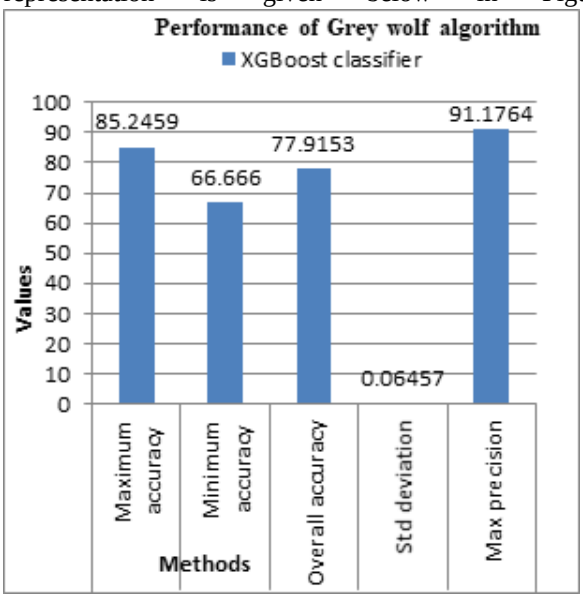


Figure.4 Graph analysis of GWO performance

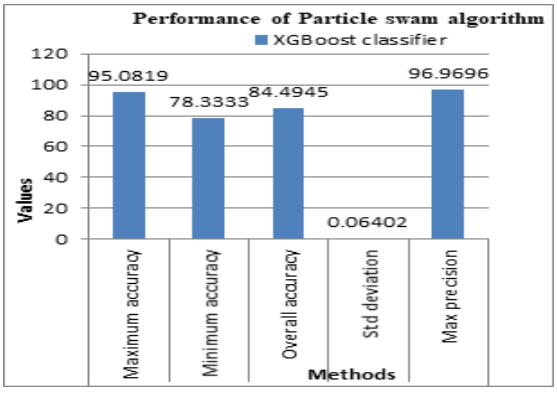
From Figure.4, it is observed that GWO has achieved greater precision value compared to GA and the tested result of PSO is given in Table.3

Performance of Feature selected by PSO algorithm		
Optimized XGBoost classifier	Maximum accuracy	95.0819
	Minimum accuracy	78.3333
	Overall accuracy	84.4945
	Std deviation	0.06402
	Max precision	96.9696

Table.3 Performance metric with PSO

From Table.3, PSO is considered to be the most efficient algorithm since the accuracy value computed from PSO is 95.0819 which is greater than remaining two algorithms and graph based representation is given in Figure. 5

Figure.5 Pictorial representation of PSO



The classification accuracy obtained from all three algorithms are represented in the following Figure.6 which gives clear picture on the efficiency of all three algorithms.

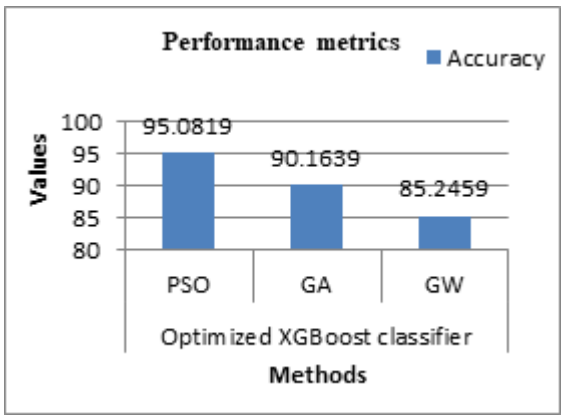


Figure.6 Accuracy analysis of classification algorithms

From Figure.6, it is clearly evident that among three meta-heuristic algorithms used for selecting best features and combined with XGBoost classifier for performing classification yields accuracy at different ranges. By comparing the accuracy values of all the methods, it is proved that PSO attains greater accuracy and precision metric compared to other algorithms.

C. Comparative analysis

The results obtained from implementation is being compared with other methods in the existing studies in order to exhibit the efficiency of the proposed system in below Table.4

Table.4 Comparative analysis of proposed accuracy with existing methods

From Table.4, it is clearly shown that the proposed work achieved higher accuracy value compared to other existing algorithms. The encryption and decryption time taken by the proposed system is being compared with the existing studies and given in below Table.5 and 6.

Key length (in bit)	**ECDHC [33]**	**MRSA [33]**	**MRSAC [33]**	**Existing approach**	**Proposed approach**
128	12	205	305	10	8.5

Table.5 Comparison of Encryption time

The time taken for encryption is being calculated and predicted that the proposed system attains minimum execution time compared to other methods. The comparison of computation time is given in below graphical view of Figure.7

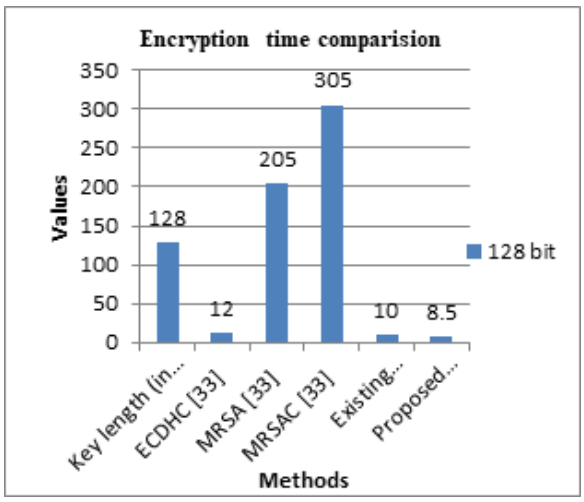


Figure.7 Graph representation of encryption time compared with existing methods

The pictorial representation of the proposed work clearly describes that the proposed system acquires minimum execution time of value 8.5 for encryption compared to other techniques.

Parameters	**LR [34]**	**KNN [34]**	**ANN [34]**	**SVM (RBF) [34]**	**SVM (Linear) [34]**	**NB [34]**	**DT [34]**	**Existing approaches**	**Proposed approach**
Accuracy (%)	88	84	82	85	87	79	78	92	95.0819

Key length (in bit)	**ECDHC [33]**	**RSA [33]**	**MRSA [33]**	**MRSAC [33]**	**Existing approach**	**Proposed approach**
128	31	188	122	188	29	15.1

Table.6 Comparison of Decryption time

The inferences made from Table.6 projects that the proposed system has achieved minimum execution time during the process of decryption of value 15.1 which is comparatively lower to other methodologies.

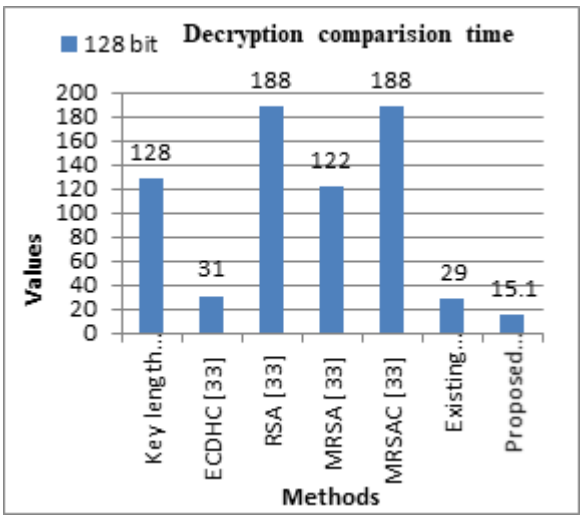


Figure.8 Time taken for decryption compared with existing methods

From the pictorial representation depicted in Figure.8, it is being observed that the proposed system achieves lower time in decryption process

Methods	Accuracy
BAT + SVM	0.4326
PUBAT + SVM	0.4711
WOA + ACNN	0.8697
Existing	0.9252
Proposed	0.95081

Table.7 Comparison of the proposed work with other techniques

The accuracy comparison made on the proposed system with other techniques have shown that the proposed implementation is efficient with value obtained as 95.081. The graphical analysis is given in below Figure.9

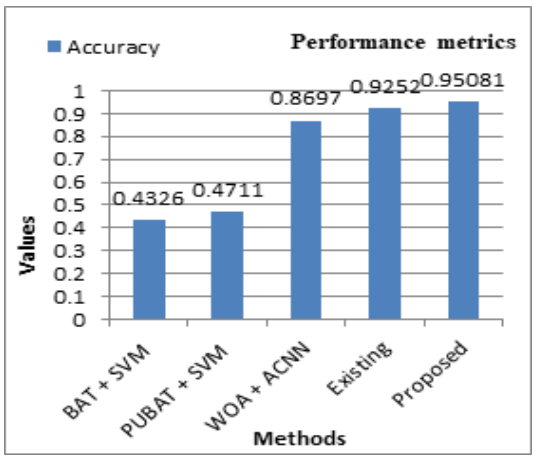


Figure.9 Accuracy comparison with other methods

From the graphical representation, it is clearly shown that the proposed framework gained better accuracy metric value when compared with other existing algorithms. The proposed algorithm accurately classifies the disease from the medical data and provides better level of security to data than existing methods.

V. Conclusion

The proposed security model utilized bit permutation for enhancing the efficiency of the modified AES algorithm. The classification process performed with meta-heuristic techniques have been evaluated. Based on the results obtained, the proposed modified AES gained efficiency due its speeder execution time and increased level of security. Among the three algorithms, PSO is considered to be efficient since it achieves greater accuracy level compared to other algorithms. Experimental results have shown that the integration of the security algorithm with classification methods have yielded better performance in diagnosing the disease based on the information retrieved from the database. The internal comparison performed have exhibited better accuracy and the comparative analysis accomplished with other methods have also gained better level of accuracy. The proposed model provides efficient performance when being implemented for medical facilities in terms of security and better diagnostic accuracy.

Declaration

Conflict of Interest: The Author reports that there is no conflict of Interest
Funding: None.
Acknowledgement: None

References

[1] A. G. de Moraes Rossetto, C. Sega, and V. R. Q. Leithardt, "An Architecture for Managing Data Privacy in Healthcare with Blockchain," *Sensors*, vol. 22, p. 8292, 2022.
[2] E. S. B. Hureib and A. A. Gutub, "Enhancing medical data security via combining elliptic curve cryptography with 1-LSB and 2-LSB image steganography," *International J Comp Sci Network Security (IJCSNS)*, vol. 20, pp. 232-241, 2020.
[3] H. M. Pandey, "Secure medical data transmission using a fusion of bit mask oriented genetic algorithm, encryption and steganography," *Future Generation Computer Systems*, vol. 111, pp. 213-225, 2020.
[4] R. Jeet, S. S. Kang, S. M. Safiul Hoque, and B. N. Dugbakie, "Secure Model for IoT Healthcare System under Encrypted Blockchain Framework," *Security and Communication Networks*, vol. 2022, 2022.
[5] S. Premkumar and J. Mohana, "A novel ECG based encryption algorithm for securing patient confidential information," *International Journal of Electrical Engineering & Technology (IJEET)*, vol. 2, pp. 35-43, 2020.
[6] J. Jayabalan and N. Jeyanthi, "Scalable blockchain model using off-chain IPFS storage for healthcare data security and privacy," *Journal of Parallel and Distributed Computing*, vol. 164, pp. 152-167, 2022.

[7] R. Shanthapriya and V. Vaithianathan, "Block-healthnet: security based healthcare system using block-chain technology," *Security Journal*, pp. 1-19, 2020.
[8] B. Raj and S. Venugopalachar, "Multi-data Multi-user End to End Encryption for Electronic Health Records Data Security in Cloud," *Wireless Personal Communications*, pp. 1-29, 2022.
[9]M. Mohammadi, R. Rawassizadeh, and A. Sheikhtaheri, "A consumer-centered security framework for sharing health data in social networks," *Journal of Information Security and Applications*, vol. 69, p. 103303, 2022.
[10] G. I. Ahmad, J. Singla, and K. J. Giri, "Security and Privacy of E-health Data," in *Multimedia Security*, ed: Springer, 2021, pp. 199-214.
[11] C. Krishnan and T. Lalitha, "Attribute-based encryption for securing healthcare data in cloud environment," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, pp. 10134-10143, 2020.
[12] G. Viswanath and P. V. Krishna, "Hybrid encryption framework for securing big data storage in multi-cloud environment," *Evolutionary Intelligence*, vol. 14, pp. 691-698, 2021.
[13] R. Denis and P. Madhubala, "Hybrid data encryption model integrating multi-objective adaptive genetic algorithm for secure medical data communication over cloud-based healthcare systems," *Multimedia Tools and Applications*, vol. 80, pp. 21165-21202, 2021.
[14] Y. M. Essa, E. E.-D. Hemdan, A. El-Mahalawy, G. Attiya, and A. El-Sayed, "IFHDS: intelligent framework for securing healthcare BigData," *Journal of medical systems*, vol. 43, pp. 1-13, 2019.
[15] M. S. Christo, P. Sarathy, and C. Priyanka, "An efficient data security in medical report using block chain technology," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, 2019, pp. 0606-0610.
[16] C. Nandni and S. Jahnavi, "Quantum Cryptography and Blockchain System: Fast and Secured Digital Communication System," in *Data Engineering and Intelligent Computing*, ed: Springer, 2021, pp. 453-462.
[17] L. M. Gupta, A. Samad, H. Garg, and K. Shah, "An Effective Meta Heuristic Based Dynamic Fine Grained Data Security Framework for Big Data," 2022.
[18] A. H. Mohsin, A. Zaidan, B. Zaidan, O. S. Albahri, A. S. Albahri, M. Alsalem, *et al.*, "Based blockchain-PSO-AES techniques in finger vein biometrics: A novel verification secure framework for patient authentication," *Computer Standards & Interfaces*, vol. 66, p. 103343, 2019.
[19] S. Doss, J. Paranthaman, S. Gopalakrishnan, A. Duraisamy, S. Pal, B. Duraisamy, *et al.*, "Memetic optimization with cryptographic encryption for secure medical data transmission in IoT-based distributed systems," *Computers, Materials & Continua*, vol. 66, pp. 1577-1594, 2021.
[20] F. Zhang, Y. Chen, W. Meng, and Q. Wu, "Hybrid encryption algorithms for medical data storage security in cloud database," *International Journal of Database Management Systems (IJDMS) Vol*, vol. 11, 2019.
[21] K. Sharma, A. Agrawal, D. Pandey, R. Khan, and S. K. Dinkar, "RSA based encryption approach for preserving confidentiality of big data," *Journal of King Saud University-Computer and Information Sciences*, 2019.
[22] R. Nidhya, S. Shanthi, and M. Kumar, "A novel encryption design for wireless body area network in remote healthcare system using enhanced RSA algorithm," in *Intelligent system design*, ed: Springer, 2021, pp. 255-263.
[23] O. N. Akande, O. C. Abikoye, A. A. Kayode, O. T. Aro, and O. R. Ogundokun, "A dynamic round triple data encryption standard cryptographic technique for data security," in *International Conference on Computational Science and Its Applications*, 2020, pp. 487-499.
[24] M. Elhoseny, K. Shankar, S. Lakshmanaprabu, A. Maseleno, and N. Arunkumar, "Hybrid optimization with cryptography encryption for medical image security in Internet of Things," *Neural computing and applications*, vol. 32, pp. 10979-10993, 2020.
[25] M. Almulhim, N. Islam, and N. Zaman, "A lightweight and secure authentication scheme for IoT based e-health applications," *International Journal of Computer Science and Network Security*, vol. 19, pp. 107-120, 2019.
[26] R. Boussada, B. Hamdane, M. E. Elhdhili, and L. A. Saidane, "Privacy-preserving aware data transmission for IoT-based e-health," *Computer Networks*, vol. 162, p. 106866, 2019.
[27] Y. Winnie, E. Umamaheswari, and D. Ajay, "Enhancing data security in IoT healthcare services using fog computing," in *2018 International Conference on Recent Trends in Advance Computing (ICRTAC)*, 2018, pp. 200-205.
[28] L. Zhang, Y. Cao, G. Zhang, Y. Huang, and C. Zheng, "A blockchain-based microgrid data disaster backup scheme in edge computing," *Security and Communication Networks*, vol. 2021, 2021.
[29] Y. Zhuang, L. R. Sheets, Y.-W. Chen, Z.-Y. Shae, J. J. Tsai, and C.-R. Shyu, "A patient-centric health information exchange framework using blockchain technology," *IEEE journal of biomedical and health informatics*, vol. 24, pp. 2169-2176, 2020.
[30] H. Qiu, M. Qiu, M. Liu, and G. Memmi, "Secure health data sharing for medical cyber-physical systems for the healthcare 4.0," *IEEE journal of biomedical and health informatics*, vol. 24, pp. 2499-2505, 2020.

Author Biographies

Uma K is a Research Scholar from the Department of Computer Science and Applications, Bangalore University, Bangalore, India. She completed her MCA master degree from Bangalore University. Her major area of research include Data Mining, Machine Learning. She has more than 10 publications in different journals.

Dr. Hanumanthappa M is a professor from the Department of Computer science and Applications, Bangalore University, Bangalore, India. He has 18+ years of Teaching, Administration and Industry Experience. Her major areas of research include Information Retrieval, Data Mining, Network Security, and Natural Language Processing. He has more than 90 publications in refereed journals. She is the reviewer of many refereed journals and also acted as advisory member for various conferences