ITC 2/52
Information Technology
and Control
Vol. 52 / No. 2 / 2023
pp. 529-540
DOI 10.5755/j01.itc.52.2.32777

Improved Smart Healthcare System of Cloud-Based IoT Framework for the Prediction of Heart Disease

Received 2022/11/15 | Accepted after revision 2023/01/07

**HOW TO CITE:** Sundararajan , S. C. M., Bharathi, G. P., Loganathan, U., Vadivel, S. (2023). Improved Smart Healthcare System of Cloud-Based IoT Framework for the Prediction of Heart Disease. *Information Technology and Control*, 52(2), 529-540. https://doi.org/10.5755/j01.itc.52.2.32777

# Improved Smart Healthcare System of Cloud-Based IoT Framework for the Prediction of Heart Disease

**Suma Christal Mary Sundararajan**

Department of Information Technology, Panimalar Engineering College (Autonomous), Poonamallee, Chennai, Tamil Nadu, India

**G. P. Bharathi**

Department of Computer and Communication Engineering, Sri Sai Ram Institute of Technology, West Tambaram, Chennai, Tamil Nadu, India

**Umasankar Loganathan**

Department of Electrical and Electronics Engineering, RMK College of Engineering and Technology, R. S. M Nagar, Puduvoyal, Chennai, Tamil Nadu, India

**Surendar Vadivel**

Department of Electrical and Electronics Engineering, Kongu Engineering College (Autonomous), Perundurai, Erode, Tamil Nadu, India

**Corresponding author:** sumachristalmaryres@outlook.com

Smart healthcare systems in the cloud-IoT framework are designed for the prediction of heart disease. This smart diagnosis improves the patient's health status and minimizes the death rate. Early prediction of heart disease may reduce the risk of patient illness and monitoring in real-time to avoid the risk. The view of existing algorithms is inaccurate in early prediction which took a lot of time for the prediction and inaccurate early prediction of heart disease. To overcome these issues, this paper proposed a sparse autoencoder with Galactic Swarm Optimization (SAE-GSO) algorithm. A sparse encoder predicts heart disease and enhances the accurate prediction, tuning the parameters of sparsity regularity in the sparse autoencoder, Galactic Swarm optimization algorithm is implemented. The proposed work enhances the prediction rate of heart diseases, minimizing the error rate, and maximizing the accuracy. The accuracy rate of the proposed work of SAE-GSO in the Cleveland Dataset produces got 92.23 %, GBT got 65.12 %, SAE got 87.34%, and NB got 83.16 %. The accuracy rate of the proposed work of SAE-GSO in the Framingham Dataset produced 92.59 %, GBT got 69.16 %, SAE got 86.25%, and NB got 82.37%.

**KEYWORDS:** Smart health, heart disease, autoencoder, Galactic Swarm Optimization, IoT.

## 1. Introduction

Smart healthcare for the prediction of a heart disease model uses sensor-based devices in the IoT platform. It accesses the data from the physical examination of patients via various sensor devices like ECG, blood pressure, body temperature, heart rate, etc. Based on these collected medical sensor data is used for the identification of diseases, decision-making process, and administration of health-related data. The automated prediction system includes the concept of IoT, Big Data, AI, and cloud networking [14]. In the cloud storage system, medical sensor signals of medical data are stored in a secure manner. This sensor data is collected dynamically on the basis of time [19]. The basic symptoms of heart disease are shortness of breath, weakness of the body, swollen feet, and arrhythmia [12].

Cardiovascular disease (CVD) is a life-threatening disease. Based on the IoT framework it captures the sensor's medical signal data and transmits it in a cloud storage environment. From the cloud storage data, a trained physician with adequate experience is able to diagnose heart disease. For the accurate prediction of heart disease, a machine learning algorithm is used and it minimizes the death rate and enhances the decision-making process clinically [7]. Mienye et al. [17] presented the prediction of heart disease by fusion of sparse autoencoder and ANN. Ebiaredoh-Mienye et al. [5] proposed a technique for predicting heart disease using a sparse autoencoder and softmax classifier. Muhammad et al. [18] described that early and on-time diagnosis of heart disease and it uses the concept of feature selection from the dataset.

The prediction of heart disease is a challenging one. Early prediction of heart disease may reduce the risk to a patient's health status. The view of existing algorithms is inaccurate in early prediction, a lot of time to predict. To overcome these issues this paper proposed a sparse autoencoder with an optimized algorithm of Galactic Swarm Optimization (SAE-GSO). For tuning the parameters of sparsity regularity in the sparse autoencoder this optimization algorithm is implemented. The proposed work enhances the prediction rate of heart diseases, minimizing the error rate, maximizing the accuracy, and minimizing the error rate.

The main contribution of the proposed work of SAE-GSO is given below:

1. In the smart healthcare system, predicting heart disease and real-time data is collected from various wearable sensor devices and the medical sensor signal data is securely stored in a cloud storage environment.
2. Effective prediction of heart disease features is selected by using SVM.
3. Based on the selected features classify the heart disease by implementing a sparse autoencoder-based optimization algorithm of Galactic Swarm Optimization.

The remaining sections of this paper are structured as follows: Section 2 discusses the related research works, Section 3 describes the prediction of heart disease using SAE-GSO, Section 4 discusses the experimented results and Section 5 concludes the proposed system with future work.

## 2. Related Works

This section discusses the literature related to smart healthcare prediction of heart disease. Cardiovascular disease (CVD) is a significantly challenging task for physicians and it increases the death rate. Therefore, machine learning and deep learning algorithms are demonstrated for the prediction of heart disease using electrocardiogram (ECG), blood pressure, heart rate, etc. and it also diagnoses heart disease at an earlier stage [8]. Nowadays, Cardiovascular disease (CVD) is the main cause of increasing death rates worldwide. The World Health Organization reports that annually 17.9 million deaths occur [29].

Nowadays, many research works have been implemented in the analysis of health-related issues based on the reports and medical records of data collected from patients. To access these medical data there are so many open sources available. Based on computer technologies, the exact diagnosis of patients and detection of disease is effective. Machine learning algorithms with artificial intelligence concept plays the main role in the diagnosis of heart disease which classifies or predict the result [25, 9]. The issues in the existing research works in the prediction of heart disease are inaccurate early predictions, and a lot of time to predict.

Dutta et al. [4] presented that the CNN model uses the two-layer technique for the prediction and analysis of disease. Maragatham et al. [13] described heart failure prediction by analyzing the big data with the LSTM concept. Onasanya et al. [21] described the concept of IoT based cloud environment for the detection of health issues and providing treatment of the patient focusing on challenges like storing information in a secure and effective manner. Pan et al. [22] presented Enhanced Deep Learning Assisted Convolutional Neural Networks for Heart Diseases Predictions on the Internet of Medical Things Platforms. Enhanced Deep Learning Assisted Convolutional Neural Networks for Heart Diseases. Predictions on the Internet of Medical Things Platforms.

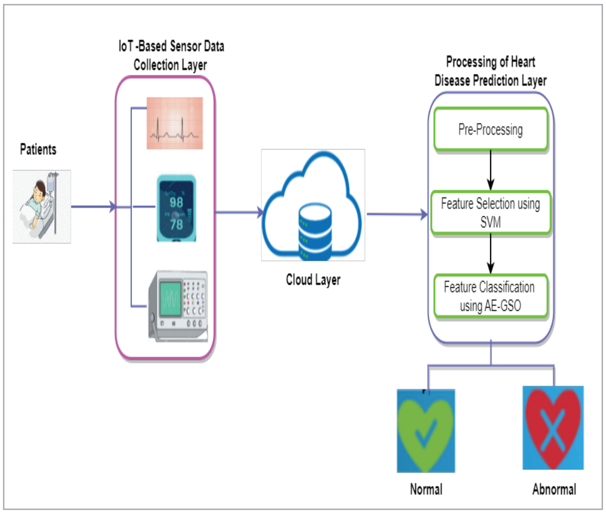
Raju et al. [2] described that IoT and Fog Computing Sectors are enabled by Cascaded Deep Learning Model for the smart prediction of heart disease. It uses optimized cascaded CNN. Nancy et al. [1] proposed the prediction of heart failure based on the IoT-Cloud-Based Deep Learning technique of Bi-LSTM (bidirectional long short-term memory). Oyeleye et al. [21] Prediction of heart rate by using autoregressive integrated moving average (ARIMA) model, linear regression, support vector regression (SVR), k-nearest neighbor (KNN) regressor, decision tree regressor, random forest regressor and long short-term memory (LSTM) recurrent neural network algorithm. Garate-Escamila et al. [8] described Classification models for heart disease prediction using feature selection and PCA. Mehmood, et al. [16] presented a Prediction of Heart Disease Using Deep Convolutional Neural Networks. Jagadeesh et al. [11] discussed Butterfly Optimized Feature Selection with Fuzzy C-Means Classifier for Thyroid Prediction. In this review of literature, we analyze many research works on the concept of the prediction of heart disease. The issues in existing works are inaccurate predictions, require more time, and require a lot of storage space for irrelevant features.

## 3. Methodology of the Proposed Framework

Smart healthcare Cloud-based IoT technology collects the various sensor signals of ECG, EEG, EMG, etc. from the patient and processes it with the proposed work of AE-GSO. The framework of AE-GSO is given in Figure 1.

Figure 1 shows the three layers of the framework namely IoT Based Data Collection Layer, the Cloud Layer, and the Processing of Heart Disease Prediction Layer.

**Figure 1**
Framework of SAE-GSO




### 3.1. IoT-Based Sensor Data Collection Layer

In the smart healthcare system, input data collected from patients through IoT devices includes the identification of patient information along with patients' past clinical history information. And also, collects medical input data like cholesterol level, heart rate, blood pressure, body temperature, blood sugar, electrocardiogram (ECG), electromyogram (EMG), and electroencephalogram (EEG) through IoT framework-based sensor devices which are attached to the body of the patient or in their clothes. The collected data is stored in the cloud for the secure monitoring of health. Table 1 shows the IoT-based secure storage of patient input information

**Table 1**
IoT-Based Secure Storage of Patient Input Information

| Identification of Patient Information | IoT-Based Patient's Clinical Data |
|---|---|
| Patient's ID | Height of Patient |
| Name of Patient | Weight of Patient |
| Patient's Gender | Cholesterol |
| Patient's Age | Drug Details |
| Address of Patient | History of Hypertension |
| Contact Number | Blood Pressure |
| | Body Temperature |
| | Heart Rate |

### 3.2. Pre-Processing

For the accurate prediction of heart disease pre-processing is required. It includes the removal of redundancy, replacement of missing values, and normalization.

#### 3.2.1. Removal of Redundancy

In the collected sensor signals, unwanted information and repeated information are removed.

#### 3.2.2. Replacement of Missing Values

In order to standardize the collection of sensor values replace the missing values with their median value. Calculating the median value by arranging the values in ascending order and identifying the median value.

#### 3.2.3. Normalization

The collected input sensor data is in an unstructured format. In normalization, the input data value is within the range of 0 and 1. This normalization helps to reduce the computing patterns of heart disease by implementing standard deviation. By processing the multiple distributions of sensor data and analysis of regression in the heart disease prediction by using:

$$A = \beta_0 + \beta_i D + \epsilon_i \text{for } i = 1,2, \ldots n. \tag{1}$$

The sample input sensor data is defined with variance and the error value is termed by $\epsilon_i$. The least square values are denoted by $\beta_0$ and $\beta_i$. From the input sensor data samples, the average values are evaluated by using standard deviation and it is defined as:

$$\mu = \frac{\sum_{i=1}^{n} D_i}{G}. \tag{2}$$

Here, $G$ represents the frequency of data, and $D_i$ is the sample input sensor data. Then the normalization is denoted by:

$$no_p = \frac{\epsilon_i^*}{\sigma_i} \tag{3}$$

$$no_p = \frac{D_i - \mu_i}{\sigma_i}. \tag{4}$$

Here, $\in_i^*$ is residual value and $\sigma_i$ is variance.

### 3.3. Feature Selection Using SVM

In the smart health care prediction of heart disease feature selection is the process of reducing the input sensor dataset for further processing and analysis to predict the heart disease based on the appropriate part of the information. This feature selection process helps to improve the performance of the prediction algorithms and it reduces the dimensionality of selecting the features. The component-wise feature vector product operator * is expressed as:

$$g * h = (g_1 h_1 \ldots . g_n h_n). \tag{5}$$

The binary feature vector $\sigma$, $\sigma \in \{0,1\}^m$ and acts as a selection of features by multiplying the component-wise featured input vector. Now the kernel function is defined as:

$$ke_\sigma\big(G_i G_s \equiv ke(\sigma * G_i, \sigma * G_s)\big). \tag{6}$$

To get the optimal hyperplane, optimization is needed.

Minimize: $r(\omega, \theta) = \frac{1}{2}\|\omega\|^2 + dp \sum_{i=1}^{j} \xi_i$

Subject to constraints:

$$\begin{gathered} h_i(\omega, \emptyset(g_i) + \theta) \geq 1 - \xi \\ \xi \geq 0, i = 1,2, \ldots, j \end{gathered}. \tag{7}$$

Here, $dp$ is the regularization parameter, by using Lagrange multiplier method Equation (7) can be rewritten as follows to get optimal vector.

$$\omega = \sum_{i=1}^{j} \alpha_i h_i \, \emptyset(g_i)$$

Subject to constraints:

$$\sum_{i=1}^{j} \alpha_i h_i = 0. \tag{8}$$

Here, $\alpha_i$ is $i = 1, 2, \ldots, j$. Substitute Equation (7) in Equation (6),

$$\begin{aligned} nd_k(g_0) &= \langle \omega, \emptyset(g_0) \rangle + \theta \\ &= \langle \textstyle\sum_{i=1}^{j} \alpha_i h_i \ \emptyset(g_i), \emptyset(g_0) \rangle + \theta \\ &= \textstyle\sum_{i=1}^{k} \alpha_i h_i \langle \emptyset(g_i), \emptyset(g_0) \rangle + \theta \end{aligned}. \tag{9}$$

Now SVM can be represented as follows:

$$sv_k(g_0) = sgn\big(md_k(g_0)\big) = \begin{cases} 1 & if\ nd_k(g_0) \geq 0 \\ -1 & if\ nd_k(g_0) < 0 \end{cases}. \tag{10}$$

Here, the support vectors are trained with training data set of non-zero value of $\alpha$.

The kernel function *ke* is used for the selection of features for the prediction of heart disease.

## 3.4. Feature Classification Using Proposed SAE-GSO Algorithm

After selecting the most relevant features from the dataset. Based on the selected features which classify the heart disease by using an autoencoder with optimized algorithm of Galactic Swarm Optimization (SAE-GSO).

The initial phase of pre-processing is done and it is discussed in Section 3.2. Steps involved in SAE-GSO is given below

**Step 1:** The sparse encoder maps the featured input data into a low dimensional space, and at the same time decoder reconstructs the featured input data. For getting an optimal solution this work implements the minimization of error. In this step bias value and weight value of matrices of encoder and decoder are employed in the model of SAE using backpropagation.

**Step 2:** Implements the optimized algorithm of GSO for optimizing the parameter of sparsity regularity.

**Step 3:** After finetuning the SAE-GSO model is employed the enhanced optimized prediction of heart disease

### 3.4.1. Autoencoder

Autoencoder is an unsupervised artificial neural network. The input selected featured vectors are designed in autoencoder. The format of autoencoders has an encoder-decoder structure [23].

Therefore, the correlation between the input features is trained and reconstructed. The operation of encoder which maps the input feature *in* to the hidden layer *hi* and it is considered as latent space representation. Similarly, the function of decoder reconstructs the latent representation as $\widehat{in}$. The process of encoding and decoding is defined as:

$$hi = \sigma(Wt.in + bi) \tag{11}$$

$$\widehat{in} = \left(wt'' hi + bi'\right). \tag{12}$$

Here, $in = (in_1, in_2, ..., in_n)$ are the featured input vector values,

$hi = (hi_1, hi_2, ..., hi_n)$ are minimized vector dimension values obtained from the hidden layer. $\widehat{in} = (\widehat{in}_1, \widehat{in}_2\ \widehat{in}_n)$ are considered as reconstructed input values. $Wt$ and $wt''$ are weight matrices and $bi$, $bi'$ bias vector values, $\sigma$ is the sigmoid activation function. It can be implemented by:

$$\sigma = \frac{1}{1+e^{-x}}. \tag{13}$$

The mean squared error function is reconstructed the error function between $hi$ and $\widehat{in}$.

$$err = \frac{1}{N}\sum_{i=1}^{N}\|\widehat{in}_i - in_i\|^2. \tag{14}$$

In training the autoencoder network, the overfitting is a challenging issue and to solve this issue by applying the weight penalty of cost function. It is defined by:

$$err = \frac{1}{N}\sum_{i=1}^{N}\|\widehat{in}_i - in_i\|^2 + \frac{\lambda}{2}(\|wt\|^2 + \|wt'\|^2). \tag{15}$$

Here, $\lambda$ is the coefficient of weight attenuation. In the autoencoder penalty weight term is introduced in hidden layer for classifying the feature. Assume that $\widehat{a}_j$ denotes the average activation function of neurons in the hidden layer and it is defined by $\widehat{a}_j = \frac{1}{M}\sum_{i=1}^{M} hid_j(in_i)$ and $\rho$ is the sparsity proportion and its value is near to 0. In order to get sparsity, kullback-Leibler (KL) divergence is used as loss function along with the limit value of $\widehat{\rho}_i = \rho$. The loss function is determined by:

$$KUL(\hat{\rho}\,||\rho) = \sum_{i=1}^{h}\rho\log\left(\frac{\rho}{\hat{\rho}_i}\right) + (1-\rho)\log\left(\frac{1-\rho}{1-\hat{\rho}_i}\right). \tag{16}$$

Here, $h$ is the neuron in the hidden layer. The loss function of sparse autoencoder is implemented in the aspects of mean squared error, sparsity regularity and weight attenuation. It is expressed as follows:

$$err = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\|\widehat{in}_i - in_i\|^2 + \frac{\lambda}{2}(\|wt\|^2 + \|wt'\|^2) + \beta KUL(\hat{\rho}\,||\rho). \tag{17}$$

Here, $\beta$ is the parameter of sparsity regularity. For enhancing the prediction of heart disease and parameter of sparsity regularity is optimized by Galactic Swarm Optimization algorithm. This proposed work of SAE-GSO produces better prediction of heart disease

### 3.4.2. Galactic Swarm Optimization

The proposed work of the smart healthcare system of cloud based IoT framework uses the sparse autoencoder with GSO, which enhances the prediction rate of heart diseases. GSO is used for optimizing the parameter of sparsity regularity hidden layers of the autoencoder network. The main objective of implementing the GSO is minimizing error rate in the aspect of Mean Square Error (MSE) and it is given by:

$$obje_{err} = argmin(MSE). \tag{18}$$

It is used to evaluate the average squares of errors in the basis of average squared error difference between the predicted result $pre_b$ and actual result $acur_b$ which is defined as:

$$MSE = \frac{1}{N}\sum_{b=i}^{N}(pre_b - acur_b)^2. \tag{19}$$

Total number of features are represented as $N$. This minimization error rate produces better accurate results in the prediction rate of heart disease with IoT assisted cloud computing. The concept behind the GSO is movement of galaxies and stars based on the influence of gravity force. Assumed the entire galaxy is considered as a point mass which fascinates the other galaxies. In the galactic swarm optimization algorithm, collection of stars in a small group is called as small galaxies and also termed as subswarms. To minimize the galaxy energy, the smaller galaxies are interlinked with one another and update their position. In addition to that, from the smaller galaxy the best star tries to communicate with the best star of the remaining other small galaxies. The best position of stars in the galaxy is represented by $posi_1$, $posi_2$, $posi_3$, ..., $posi_m$ and its velocities are represented as $\vartheta_1$, $\vartheta_2$, $\vartheta_3$, $\vartheta_4$, $\vartheta_5$. To form the super swarm by grouping the star and the best star $b$ with its velocity $vel$. To minimize the energy of the galaxy system, it updates the velocity with its position of superswarm galaxy and it maintains the stabilization.

Assuming the cluster $cltr$ of sub swarm has $r$ tuples include elements of $e_{i,j} \in R^m$, which contains $n$ partitions and size of each partition is $m$ and its formulated as

$$Q_i \subset cltr;\ \forall_i[1,m], d_j \subset Q_j;\ \forall_j[1,t], Q_i \cap Q_j = \emptyset;\ \ if\ i \neq j, \cup_{i=1}^{m} Q_i = cltr. \tag{20}$$

In the implementation of GSO contains two levels of process namely sub warms and super warms. It contains m sub warms and each sub warm $cltr_i$ contains best star which is interrelated with one another. Its updated velocity of best star is denoted by $velbe_i$. In a galaxy sub warm $cltr_i$ which explores all search space with its updated position and velocity of every swarm in the galaxy is represented as:

$$d_{i,j} = \left(\vartheta_{i,j} + d_{i,j}\right) \tag{21}$$

$$\vartheta_{i,j} = \left\{\left(\omega_i \times \vartheta_{i,j} + cn_1 \times rand_1\right) \times \left(velbe_{i,j} - d_{i,j}\right) + dn_2 \times rand_2 \times \left(cltr_{i,j} - d_{i,j}\right)\right\}. \tag{22}$$

Here, the random values are $rand_1$ and $rand_2$ with its initial weight is denoted as $\varpi_i$ and its updated values are:

$$rand_i = \cup\ (0,1) \tag{23}$$

$$\omega_i = 1 - \frac{itera}{itera_{max}+1}. \tag{24}$$

Here, $rand_i$ random number is range between 0 to 1 and its current iteration is $itera$ which is varied from [1, $itera_{max}$]. The formation of superclusters or superwarms are implemented when it is assisted by best subwarms of next phase clustering. The velocity and position of sub warms in the super warm are updated by using:

$$Pos_{i,j} = vl_{i,j} + Pos_{i,j} \tag{25}$$

$$vl_{i,j} = \left\{\left(\omega_i \times vl_{i,j} + dn_1 \times rnd_1\right) \times \left(velbe_{i,j} - g_i\right) + dn_2 \times rand_2 \times (Pos - Pos_i)\right\}. \tag{26}$$

Here, the best star in the galaxy is denoted Pos which is updated its velocity by using the Equation (26). The flowchart of GSO is given in Figure 2.

## 4. Results and Discussion

This proposed work of the smart healthcare system of cloud based IoT framework of SAE-GSO is used to predict the heart disease model was implemented in Python. This proposed method is compared with Gradient Boosted Tree (GBT) [3], Sparse Autoencoder (SAE) [10], Naïve Bayes classifier (NB) [24].

**Figure 2**
Flowchart of GSO

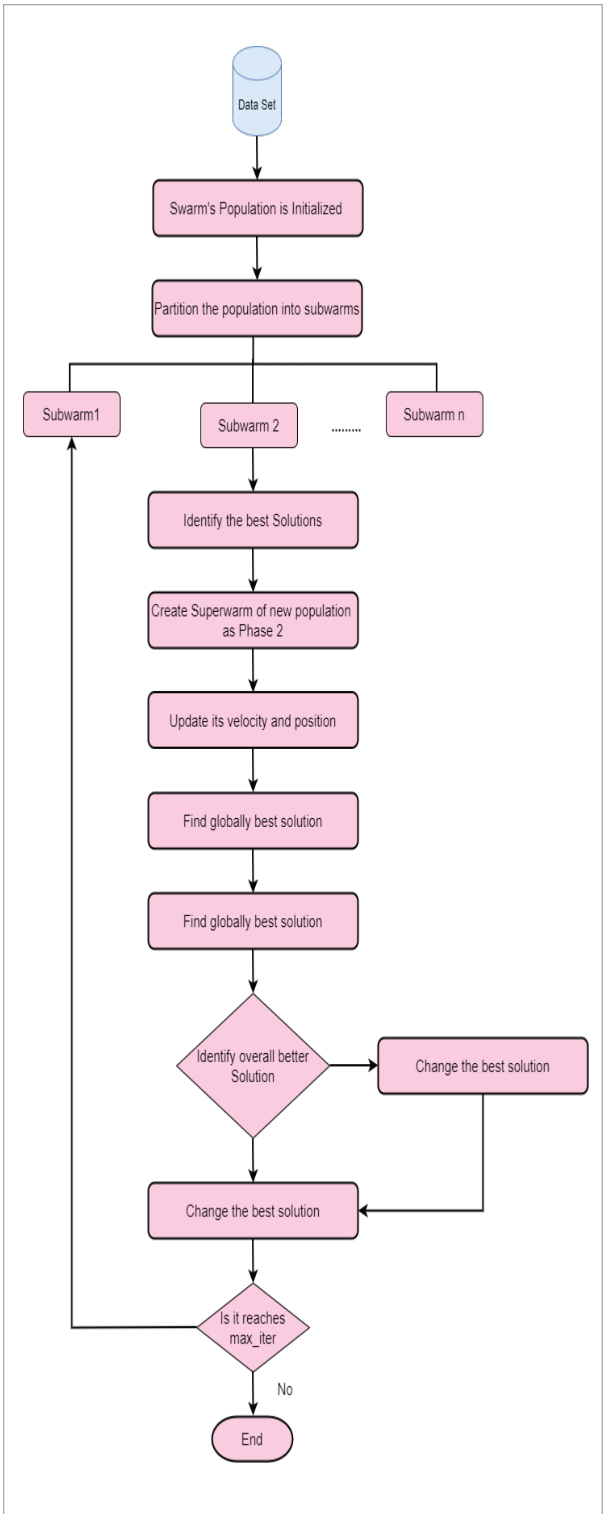



**Table 2**
Description of features in the Cleveland dataset

| Features | Description |
|---|---|
| f1-Age | Patient's Age in years |
| f2-Sex | Female or Male |
| f3-Chol | Measure of serum cholesterol |
| f4-Cp | Types of chest pain |
| f5-Trestbps | Resting blood pressure |
| f6-Fbs | fasting blood sugar |
| f7-Restecg | Resting measure of electrocardiographic |
| f8-Thalach | maximum heart rate |
| f9-Exang | Exercise caused angina |
| f10-Oldpeak ST | ST depression caused by exercise relative to rest |
| f11-Slope | Slope of the peak exercise ST segment |
| f12-Ca | Number of major vessels of colored by Flourosopy |
| f13-Thal | Type of Defect |

**Table 3**
Description of Framingham dataset

| Features | Description |
|---|---|
| f1-Sex | Gender of patient |
| f2 Age | Age of Patient |
| f3- Education | Educational level of patient (1 = high school, 2 = GED certificate, 3 = vocational training,4 = college degree) |
| f4-Currentsmoker | Whether the patient smoker or not |
| f5-cigsPerDay | The average number of cigarettes the patient smokes per day |
| f6-prevalentStroke | Whether the individual previously had a stroke or not |
| f7-Diabetes | Whether the individual is diabetic |
| f8-totChol | Total cholesterol level |
| f9-sysBP | Systolic blood pressure |
| f10-diaBP | Diastolic blood pressure |
| f11-BMI | Body mass index |
| f12-heartrate | The patient's heart rate |
| f13-Glucose | Glucose level |
| f14-Target variable (TenYearCHD) | Whether or not the patient has a ten-year risk of coronary heart disease |

## 4.1. Data Set Description

The dataset used in the smart healthcare system of cloud based IoT framework for the prediction of heart disease are Cleveland dataset from the UCI repository [26] and Framingham dataset [6]. Data collected from physical examination of patients via various sensor devices like ECG, blood pressure, body temperature, heart rate etc. It consists of 303 and 294 records with 13 features. Meanwhile, the Framingham dataset contains 4238 samples with 14 features. Table 2 is the description of the Cleveland dataset. Table 3 shows that description of the Framingham dataset.

## 4.2. Performance Metric Measures

These parametric metric measures are computed and predicted the heart disease based on the Cleveland Dataset and Framingham dataset using proposed work SAE-GSO.

$$Sensitivity = \frac{TP}{TP+FN} \tag{27}$$

$$Specificty = \frac{TN}{TN+FP} \tag{28}$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{29}$$

$$miss\ rate(FPR) = \frac{FN}{TP+FN} \tag{30}$$

$$fall\ out(FNR) = \frac{FP}{TN+FP} \tag{31}$$

$$F-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{32}$$

**MSE**

The mean squared error (MSE) calculate the average of the squares of the differences between the predicted values and actual values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{pi} - y_{ai}\right)^2 \tag{33}$$

**RMSE**

It is similar to MSE but for compute this by root of MSE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{pi} - y_{ai}\right)^2} \tag{34}$$

The performance of various feature selection algorithms in Cleveland dataset and Framingham dataset is implemented based on Logistic Regression [27], Random Forest (RF) [30], hybrid algorithm of SVM with Genetic algorithm (SVM-GA) [28]. Table 4 shows that comparison of feature selection in the Cleveland dataset.

**Table 4**

Comparison of Feature Selection in Cleveland dataset

| Feature Selection | Number of Features | Selected Features |
|---|---|---|
| Logistic Regression | 13 | f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12,f13 |
| RF | 11 | f1,f2,f3,f6,f7,f8,f9,f10,f11,f12,f13 |
| SVM-GA | 7 | f2,f3,f4,f5,f6,f7,f8 |
| SVM-SAE-GSO (proposed) | 5 | f3,f4,f6,f7,f8 |

In the observation of Table 4, it seems that selection of features from the dataset produces prominent results by the proposed algorithm of SVM-SAE-GSO. Table 5 shows the statistical metric measures of the Framingham dataset.

**Table 5**

Comparison of Feature Selection in Framingham Dataset

| Feature Selection | Number of Features | Selected Features |
|---|---|---|
| Logistic Regression | 13 | f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12,f13 |
| RF | 10 | f1,f2,f3,f6,f7,f8,f9,f10,f11,f12 |
| SVM-GA | 8 | f2,f3,f4,f5,f6,f7,f8,f11 |
| SVM-SAE-GSO | 5 | f3,f4,f6,f7,f9 |

In the observation of Table 5, it seems that selection of features from the dataset produces prominent results by the proposed algorithm of SVM-SAE-GSO. Table 6 shows the analysis of various metric measures by applying the Cleveland dataset.

**Table 6**
Statistical Metric Measures of Cleveland Dataset

| Algorithm | Sensitivity | Specificity | F1-Score | Miss Rate | Fall\out |
|---|---|---|---|---|---|
| **GBT** | 85.1 | 73.7 | 82.6 | 77.6 | 72.4 |
| **SAE** | 65.9 | 67.3 | 61.9 | 59.8 | 65.4 |
| **NB** | 87.9 | 86.5 | 83.1 | 55.3 | 61.9 |
| **SAE-GSO** | 94.8 | 95.2 | 94.7 | 42.6 | 53.2 |

Table 6 shows the performance of statistical metric measures of sensitivity, specificity, F1-Score, miss rate, and fall out. The GBT algorithm produces 85.1% in sensitivity, 73.7% in specificity, 82.6% in F1-Score, 77.6 % in miss rate and 72.4% in fall out. Meanwhile for SAE the parametric metric measures of sensitivity got 65.9 %, for specificity 67.3 %, F1-score 61.9 %, Miss Rate 59.8%, and Fallout 65.4 %. For the NB algorithm the parametric metric measures of sensitivity got 87.9 %, for specificity 86.5 %, F1-score 83.1 %, Miss Rate 55.3%, and Fall out 61.9%. Proposed work of SAE-GSO produces the prominent results in the aspects of sensitivity 94.8%, specificity 95.2%, F1-score 94.7 %, Miss Rate 42.6%, and fall out 53.2%.

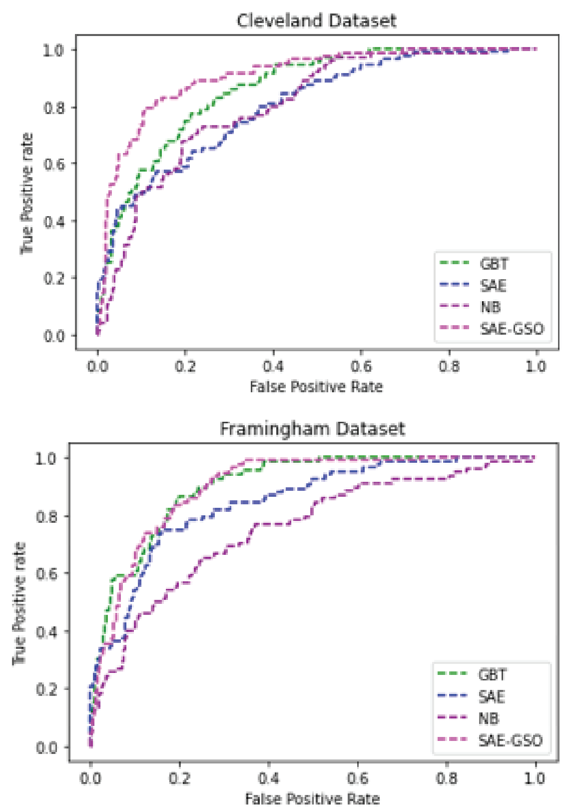
Table 7 shows the analysis of various metric measures by applying the Cleveland dataset.

Table 6 shows the performance of statistical metric measures of sensitivity, specificity, F1-Score, miss rate, and fall out. The GBT algorithm produces 81.8% in sensitivity, 77.3 % in specificity, 79.6 % in F1-Score, 75.4 % in miss rate and 76.8% in fall out. Meanwhile for SAE the parametric metric measures of sensitivity got 68.2 %, for specificity 64.6 %, F1-score 63.9 %, Miss Rate 54.5 %, and Fallout 62.3 %. For the NB algorithm, the parametric metric measures of sensitivity got 84.3 %, for specificity 85.7 %, F1-score 84.6%, Miss Rate 53.8 %, and Fall out 62.6 %. Proposed work of SAE-GSO produces the prominent results in the aspects of sensitivity 95.9 %, specificity 94.7 %, F1-score 93.2 %, Miss Rate 49.4 %, and Fallout 55.2%. Figure 3 shows the ROC of comparison of various algorithms in two different datasets.

**Table 7**
Statistical Metric Measures of Framingham Dataset

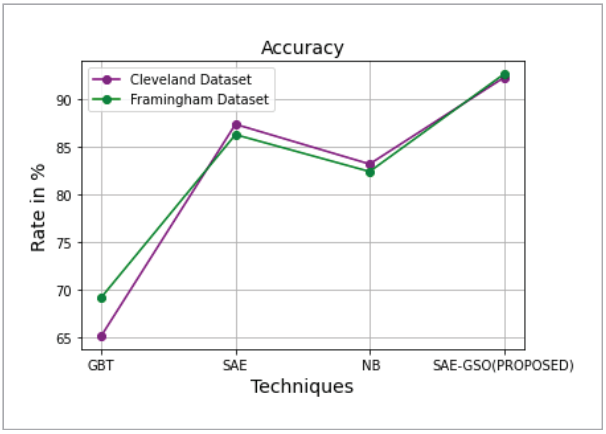
| Algorithm | Sensitivity | Specificity | F1-Score | Miss Rate | Fall\out |
|---|---|---|---|---|---|
| **GBT** | 81.8 | 77.3 | 79.6 | 75.4 | 76.8 |
| **SAE** | 68.2 | 64.6 | 63.9 | 54.5 | 62.3 |
| **NB** | 84.3 | 85.7 | 84.6 | 53.8 | 62.6 |
| **SAE-GSO** | 95.9 | 94.7 | 93.2 | 49.4 | 55.2 |

**Figure 3**
ROC




In the observation of Figure 3, ROC of various algorithms like GBT, SAE, NB and SAE-GSO. Our proposed work of SAE-GSO produces a better outcome compared with other techniques. The ROC curve plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 – Specificity) at various thresh-

**Figure 4**
Accuracy




old values. SAE-GSO is capable of producing the best classifier in the prediction of heart disease in the both datasets. Figure 5 shows the accuracy rate of various algorithms in the both datasets of Cleveland Dataset, Framingham Dataset.

Figure 4 shows that the accuracy rate of proposed work of SAE-GSO in the Cleveland Dataset produces got 92.23 %, GBT got 65.12 %, SAE got 87.34%, NB got 83.16 %. The accuracy rate of proposed work of SAE-GSO in the Framingham Dataset produced 92.59 %, GBT got 69.16 %, SAE got 86.25%, NB got 82.37%. Figure 5 shows the analysis of error rate.

Figure 5 is an analysis of the error rate of MSE and RMSE using various algorithms of GBT, SAE, NB and SAE-GSO for the prediction of heart disease. Our proposed work SAE-GSO produces a minimum error rate in the prediction of heart disease when compared with other techniques and also it is implemented in two different datasets. Figure 6 shows that computation time.

Figure 6 observes that implementation of various techniques in the prediction of heart disease our proposed work SAE-GSO requires minimum time in both datasets of Cleveland Dataset and Framingham Dataset. From the above analysis of performance of proposed work produces prominent results in the aspects of accuracy, sensitivity, specificity and F1-score, Miss rate, Fall out and predict the heart disease accurately in a minimum computation time and minimum error rate.

**Figure 5**
Error Rate

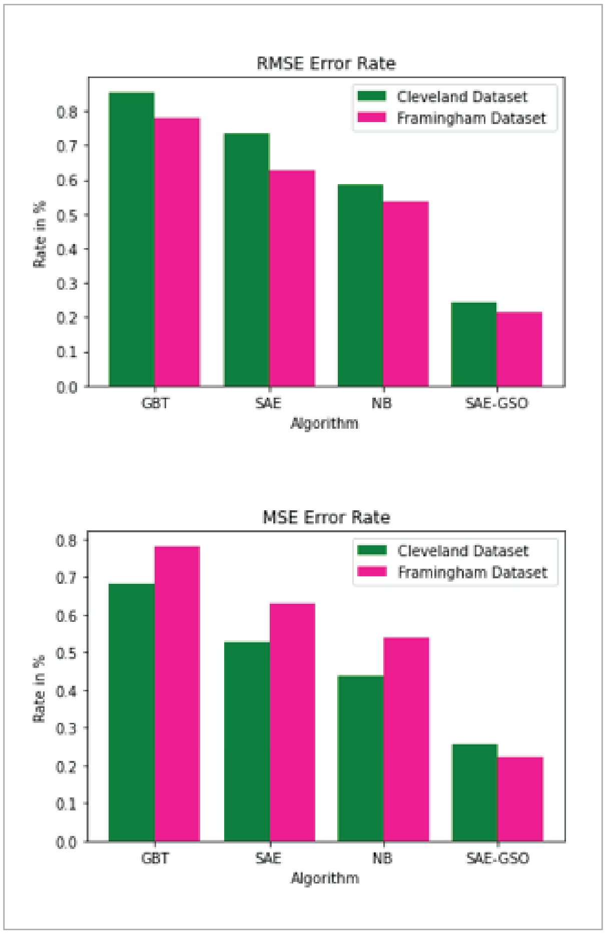
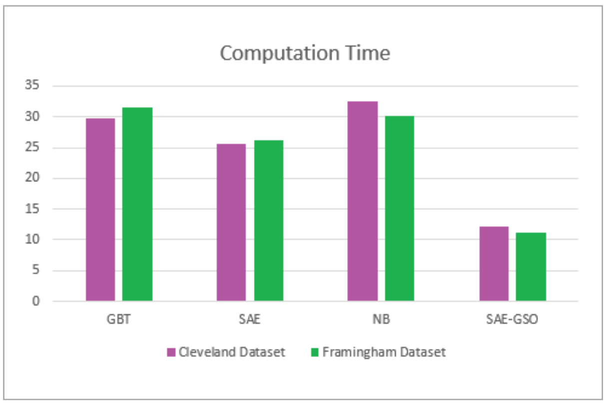



**Figure 6**
Computation Time



## 5. Conclusion

This paper presented a smart health care system for predicting heart disease based on the IoT framework. For the effective prediction of heart disease, feature selection is implemented by applying SVM. Based on the selection of features, proposed work uses sparse autoencoder and parameter of sparsity regularity is optimized by Galactic Swarm Optimization algorithm. This optimization algorithm is enhancing the prediction of heart disease in an effective way. This work uses two different types of datasets namely Cleveland Dataset and Framingham Dataset. The accuracy rate of proposed work of SAE-GSO in the Cleveland Dataset produces got 92.23 %, GBT got 65.12 %, SAE got 87.34%, NB got 83.16 %. The accuracy rate of proposed work of SAE-GSO in the Framingham Dataset produced 92.59 %, GBT got 69.16 %, SAE got 86.25%, NB got 82.37%. In the future, this work may be implemented by fuzzy based algorithms.

## References


1. Angel, N. A., Dakshanamoorthy, R., Durai, R., Vincent, P. M., Kathiravan, S., Daniel, G., Gutierrez, R. IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning. Electronics, 2022, 11, 2292. https://doi.org/10.3390/electronics11152292
2. Butchi, R. K., Suresh, D., Ankit, V., Vidyarthi, V., Gupta, M. N. S. S. V. K. R., Baseem, K. Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model. Computational Intelligence and Neuroscience, 2022, 2022, 1070697. https://doi.org/10.1155/2022/1070697
3. Danish, H., Syed, S., Ullah, J. I., Hussain, S., Anwar ul Hassan, C., Umar, F. Hybrid Feature Selection Method Based on Genetic Algorithm for the Diagnosis of Coronary Heart Disease. Journal of Information and Communication Convergence Engineering, 2022, 20(1), 31-40.
4. Dutta, A., Batabyal, T., Basu, M., Acton, S. T. An Efficient Convolutional Neural Network for Coronary Heart Disease Prediction. Expert Systems with Applications, 2020, 159, 113408. https://doi.org/10.1016/j.eswa.2020.113408
5. Ebiaredoh Mienye, S. A., Esenogho, E., Swart, T. G. Integrating Enhanced Sparse Autoencoder-based Artificial Neural Network Technique and Softmax Regression for Medical Diagnosis. Electronics, 2020, 9, 1963. Framingham Heart Study Dataset. Available online: https://kaggle.com/amanajmera1/framingham-heart-study-dataset. https://doi.org/10.3390/electronics9111963
6. Fitriyani, N. L., Syafrudin, M., Alfian, G., Rhee, J. HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System. IEEE Access, 2020, 8, 133034-133050. https://doi.org/10.1109/ACCESS.2020.3010511
7. Garate-Escamila, A. K., Hajjam, E., Hassani, A., Andrés, E. Classification Models for Heart Disease Prediction Using Feature Selection and PCA. Informatics in Medicine Unlocked, 2020, 19, 100330. https://doi.org/10.1016/j.imu.2020.100330
8. Hastie, T., Tibshirani, R., Friedman, J. An Element of Statistical Learning, Data Mining, Inference, and Prediction. Springer, Cham, Switzerland, 2020.
9. Ibomoiye, D., Mienye, Sun, Y. Improved Heart Disease Prediction Using Particle Swarm Optimization Based Stacked Sparse Autoencoder. Electronics, 2021, 10, 2347. https://doi.org/10.3390/electronics10192347
10. Jagadeesh Kumar, S. J. K., Parthasarathi, P., Mehedi, M., Jehad Al-Amri, F., Mohamed, A., Abouhawwash, Butterfly Optimized Feature Selection with Fuzzy C-Means Classifier for Thyroid Prediction. Intelligent Automation & Soft Computing, 2022, 35(3), 2909-2924. https://doi.org/10.32604/iasc.2023.030335
11. Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., Saboor, A. Heart Disease Identification Method using Machine Learning Classification in E-Healthcare. IEEE Access, 2020, 8, 107562-107582. https://doi.org/10.1109/ACCESS.2020.3001149
12. Maragatham, G., Devi, S. LSTM Model for Prediction of Heart Failure in Big Data. Journal of Medical Systems, 2019, 19. https://doi.org/10.1007/s10916-019-1243-3
13. Manimurugan, S. Two-Stage Classification Model for the Prediction of Heart Disease Using IoMT and Artificial Intelligence. Sensors, 2022, 22(2), 476. https://doi.org/10.3390/s22020476
14. Md Ahsan, M., Siddique, Z. Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review, 2021. arXiv: 2112.06459v1, 2021

15. Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., Masood, M. Prediction of Heart Disease Using Deep Convolutional Neural Networks. Arabian Journal for Science and Engineering, 2021, 46, 3409-3422. https://doi.org/10.1007/s13369-020-05105-1

16. Mienye, I. D., Sun, Y., Wang, Z. Improved Sparse Autoencoder Based Artificial Neural Network Approach for Prediction of Heart Disease. Informatics in Medicine Unlocked, 2020, 18, 100307. https://doi.org/10.1016/j.imu.2020.100307

17. Muhammad, Y., Tahir, M., Hayat, M., Chong, K. T. Early and Accurate Detection and Diagnosis of Heart Disease Using Intelligent Computational Model. Scientific Reports, 2020, 10(1), 1-17. https://doi.org/10.1038/s41598-020-76635-9

18. Nascimento L. M, Sensors and Systems for Physical Rehabilitation and Health Monitoring - A Review. Sensors, 2020, 20(15), 4063. https://doi.org/10.3390/s20154063

19. Onasanya, A., Elshakankiri, M. Smart Integrated IoT Healthcare System for Heart Care. Wireless Network, 2019, 27(6), 4297-312. https://doi.org/10.1007/s11276-018-01932-1

20. Oyeleye, M., Chen, T., Titarenko, S., Antoniou, G. A Predictive Analysis of Heart Rates Using Machine Learning Techniques. International Journal of Environmental Research and Public Health, 2022, 19, 2417. https://doi.org/10.3390/ijerph19042417

21. Pan, Y., Fu, M., Cheng, B., Tao, X., Guo, J. Enhanced Deep Learning Assisted Convolutional Neural Networks for Heart Diseases Predictions on the Internet of Medical Things Platforms. IEEE Access, 2020, 8, 189503-189512. https://doi.org/10.1109/ACCESS.2020.3026214

22. Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., Wang, R. Development and Application of a Deep Learning-Based Sparse Autoencoder Framework for Structural Damage Identification. Structural Health Monitoring, 2019, 18, 103-122. https://doi.org/10.1177/1475921718800363

23. Selvan, S., Thangaraj, S., J. J., Samson Isaac, J., Benil, T., Muthulakshmi, K., Almoallim, H. S., Alharbi, S. A., Kumar, R. R., Thimothy, S. P. An Image Processing Approach for Detection of Prenatal Heart Disease. BioMed Research International, 2022, Article ID 2003184. https://doi.org/10.1155/2022/2003184

24. Shalev-Shwartz S., Ben-David, S. Understanding Machine Learning, From Theory to Algorithms. Cambridge University Press, Cambridge, UK, 2020.

25. UCI Machine Learning Repository. Uci.edu. Available online: http://archive.ics.uci.edu/ml.

26. Usha, S., Kanchana, S. Predicting Heart Disease Using Feature Selection Techniques Based on Data Driven Approach. Webology, 2021, 18(4).

27. Usha, S., Kanchana, S. Predicting Heart Disease Using Feature Selection Techniques Based on Data Driven Approach, Webology, 2021, 18(4).

28. Wiharto, E. S., Setyawan, S., Putra, B. P. E. Hybrid Feature Selection Method Based on Genetic Algorithm for the Diagnosis of Coronary Heart Disease. Journal of Information and Communication Convergence Engineering, 2022, 20(1), 31-40.

29. World Health Organization. Cardiovascular Diseases. WHO, Geneva, Switzerland, 2020. https://www.who.int/healthtopics/cardiovascular-diseases/#tab_1.

30. Yadav, D., Pal, S. Prediction of Heart Disease Using Feature Selection and Random Forest Ensemble Method. International Journal for Pharmaceutical Research Scholars, 2020, 12, 56-66. https://doi.org/10.31838/ijpr/2020.12.04.013