Journal of Current Science and Technology, May-September 2023 Vol. 13 No. 2, pp. 392-411
Copyright ©2018-2023, Rangsit University ISSN 2630-0656 (Online)

Cite this article: Gopu, K. L. K., & Kannan, S. (2023, May). Deep Graph Neural Network with Fish-Inspired Task Allocation Algorithm for Heart Disease Diagnosis. *Journal of Current Science and Technology, 13*(2), 392-411. https://doi.org/10.59796/jcst.V13N2.2023.1753

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Deep Graph Neural Network with Fish-Inspired Task Allocation Algorithm for Heart Disease Diagnosis

Krishna Lava Kumar Gopu[1*] and Suthendran Kannan[2]

[1]Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad,
Telangana, 501401, India
[2]Department of Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil,
Srivilliputhur, Tamil Nadu 626126, India

[1*]Corresponding author; E-mail: lavakumar.gopu@gmail.com
[2]Email: k.suthendran@klu.ac.in

Received 11 April 2022; Revised 28 March 2023; Accepted 30 March 2023;
Published online 15 July 2023

Abstract

Heart disease is a very hazardous disease and many people suffered from this disease globally. The major aim is “to diagnose the heart disease with higher accuracy through decreasing error rate including computational complexity”. The existing techniques did not proffer adequate accuracy and also increased the error rate. Therefore, a Deep Graph Neural Network with Fish-Inspired Task Allocation Algorithm is proposed in this manuscript for categorizing heart disease diagnosis (DGNN-FITA-HDD). Synthetic Minority Oversampling and standard scalar strategies are utilized for pre-processing process. The pre-processed output is given to feature selection process. Two-Stage Feature Selection method selects the most important features from pre-processing output. Extracted features are transferred to Deep Graph Neural Network (DGNN) for categorizing presence and absence of heart disease. DGNN does not expose any adoption of optimization strategies for calculating the optimum parameters to assure accurate prediction. Fish-Inspired Task Allocation approach is proposed for optimizing the weight parameters of DGNN. The proposed approach is executed at MATLAB. The performance of algorithm is analyzed with/without feature selection method. By this, the proposed DGNN-FITA-HDD method attains higher accuracy with feature selection of 13.41%, 18.53%, 10.38% and 9.31% and without feature selection attains 6.5%, 8.64%, 4.39%, and 10.28% compared with existing methods, like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively.

Keywords: *deep graph neural network; feature selection; heart disease; machine learning; SMOT*

1. Introduction

Heart disease is the most dangerous health issue and several people have been affected through this disease (Reddy et al., 2020). Heart disease causes common symptoms, like shortness of breath, weakness and swelling of the legs (Abdel-Basset, Gamal, Manogaran, Son, & Long, 2020). Researchers try to proficient technique to the diagnosis of heart disease, because existing methods are not much effective in earlier detection (Thanga Selvi, & Muthulakshmi, 2021). Heart disease is very difficult to diagnose and treat in the absence of modern technology and medical professionals (Li et al., 2020). Effectual diagnosis including better treatment can save many people lives (Nourmohammadi-Khiarak et al., 2020). As per the European Society of Cardiology, 26 million people are infected along Heart Disease and 3.6 million are diagnosed every year (Saqlain et al., 2019). The Heart Disease diagnosis is traditionally made through patient’s medical history analysis, physical examination report, and related symptoms. But outcomes acquired from identification method are not accurate for identifying Heart Disease patients (Khan, & Algarni, 2020; Einstein et al., 2021; Liu, Tian, Zhao, Huang, & Wang, 2020; Shajin, &Rajesh, 2022). This is expensive as well as computational complex to analyze (Shaji, 2019; Rajesh, Shajin, Mouli Chandra, & Kommula, 2021; Shajin, Rajesh, & Raja, 2021). To handle these issues, a non-invasive detection schemeis developed depending on deep learning (DL) classifiers (Gokulnath, & Shantharajah, 2019; Rajesh, Muthubalaji, Srinivasan, & Shajin, 2022). The deep learning predictive methods require appropriate data for testing and training. Deep learning can enhance the effectiveness of the model if a balanced database is used for training with testing model (Khan, 2020; Wang et al., 2020; Mansour et al., 2021). The model predictive capacities can be improved with respect to the relevant features from data. So, data balancing and feature selection are considerably significant for the improvement of method performance (Tesson, Butow, Marshall, Fonagy, & Kasparian, 2021).

The main challenge of Heart Disease is its diagnosis. There are tools can predict Heart Disease, but they are expensive or ineffective for estimating risk for Heart Disease in humans. Early detection only lessens the mortality rate. However, in all cases the patient cannot be accurately monitored every day and a patient is not available for consultation 24 hours a day. Existing studies explores the use of heart disease with a computer-aided diagnosis (CADx) system to improve the heart disease prediction performance. The existing studies consist of numerous feature extractions with classification approaches, but that approaches contains higher over fitting problem resulting lower detection accuracy. Heart Disease is one of the lives threatening which directly affects human life. Significant heart disease task is properly detected heart disease phases for prevention and cure of illness. Numerous deep learning techniques were suggested for detecting heart disease. But, existing technique does not offer the sufficient accuracy and rises the computation time when identifying heart disease that is motivated to do this work.

1.1 Literature Survey

Various research works related to heart disease identification were presented at the literature previously, few recent research works are described below,

Bakhsh (2021) have presented an Enhanced Deep Genetic Algorithm with Adaptive Harris Hawk optimization algorithm (EDGA-AHHO) to detect and classify the heart diseases. For feature selection process, stochastic gradient boosting based recursive feature elimination methodology was presented, which easily identify the people with heart disease. Moreover, the EDGA-AHHO method analyses heart disease in the publicly accessible dataset. The EDGA-AHHO model was effectively identifies the heart diseases, but the classification accuracy was low conversely with higher sensitivity.

Sharma et al., (2020) have presented an Extreme Gradient Boosting (XGB) with Modified Artificial Plant Optimization (MAPO) approach for heart disease prediction in terms of heart rate. The MAPO mechanism was utilized for performing feature selection process. Also, XGB-MAPO model analyzed the heart disease on fingertip video dataset, also detect the Coronary Heart Disease in every patient was presented in the dataset. Thus, the XGB-MAPO method was effectively predict the heart rate with better performance in terms of Standard Error and Pearson correlation but the accuracy of the XGB-MAPO method was low.

Reddy et al., (2020) have presented hybrid genetic approach along fuzzy logic classifier for HD detection. The intention of presented model was predicting the HD that helps medical practitioners for identifying initial stage. Where, rough sets base HD feature selection mode and fuzzy rule base classification model were used. The developed rules from fuzzy classifiers were optimized through utilizing adaptive genetic approach. It offers lower accuracy with better sensitivity.

Ghosh et al., (2021) have presented proficient cardiovascular prediction disease utilizing machine learning approaches along relief and Least Absolute Shrinkage and Selection Operator (LASSO) feature selection methods. Random Forest Bagging Method (RFBM) was suggested by incorporating traditional classifiers along bagging and boosting techniques employed at the training procedure. It offers better precision along lesser accuracy.

Nilashi et al., (2020) have presented coronary HD identification utilizing self-organizing map, fuzzy support vector machine along incremental updates. The suggested technique in terms of principal module analysis, fuzzy support vector machine (FSVM), self-organizing map, 2 imputation strategies for missing value accusation. The incremental principal component analysis with FSVM was applied to decrease the computation time. It provides lesser precision along higher accuracy.

Li et al., (2020) have presented HD identification utilizing machine learning categorization at e-healthcare. Where, conditional mutual information feature selection algorithm was employed for features selection to increase categorizations accuracy, also decrease execution time. The leave one subject out cross-validation was employed for learning best practices of method assessment and hyper parameter tuning. It provides lesser accuracy with lower error rate.

Nourmohammadi-Khiarak et al., (2020) have presented hybrid technique for heart disease diagnosis using optimization approach in feature selection. Where, imperialist competitive along meta-heuristic algorithm was presented to select vital features of Heart Disease. The presented approach provides optimal response for feature selection. K-nearest neighbour approach was deemed for the categorizations. It provides higher accuracy with lower sensitivity.

2. Objectives

A Deep Graph Neural Network with Fish-Inspired Task Allocation Algorithm for classifying heart disease with high accuracy and low error rate. Here, the input data are pre-processed and features are selected for enhancing the categorization. Thus, the proposed DGNN-FITA-HDD approach classifies the heart diseases with high performance. The major contributions of this manuscript is given below,

- A Deep Graph Neural Network along Fish-Inspired Task Allocation Algorithm is proposed for classifying heart disease with high accuracy and low error rate.
- Initially, Synthetic Minority Oversampling Technique with standard scalar methods (SMOT-SS) (Rani, Kumar, Ahmed, & Jain 2021) is used for pre-processing process.
- Then pre-processed output is given to the feature selection procedure. Here, Two-Stage Feature Selection method (Niu, Wang, Lu, Yang, & Du, 2020) selects the most important features from the pre-processing output.
- At last, Deep Graph Neural Network receives the extracted features (Guo et al., 2021) for classifying absence and presence of heart diseases.
- In general, DGNN does not reveal any adoption of optimization methods for scaling optimum parameters with ensuring correct prediction.
- Therefore, a Fish-Inspired Task Allocation (Alhaqbani, Kurdi, & Youcef-Toumi, 2020) for optimizing the DGNN weight parameters.
- The proposed algorithm is activated at MATLAB software and efficacy of technique is examined under certain performances metrics, such as f-measure, accuracy, error rate, recall, precision, specificity, sensitivity, computational complexity, ROC.
- The performance of technique is comparing to the existing EDGA-AHHO-HDD (Bakhsh, 2021), XGB-MAPO-HDD (Sharma et al., 2020), AGAFL-HDD (Reddy et al., 2020) and RFBM-HDD (Ghosh et al., 2021) methods respectively.

The remaining manuscript is organized as: segment 3 explains the proposed algorithm, segment 4 demonstrates outcomes, and segment 5 concludes the manuscript.

3. Proposed DGNN-FITA-HDD Methodology

Heart disease is a major health problem and affects many people around the world. The medical practitioners can gain valuable insights about patients using machine learning techniques that help to provide proper treatment for patients.

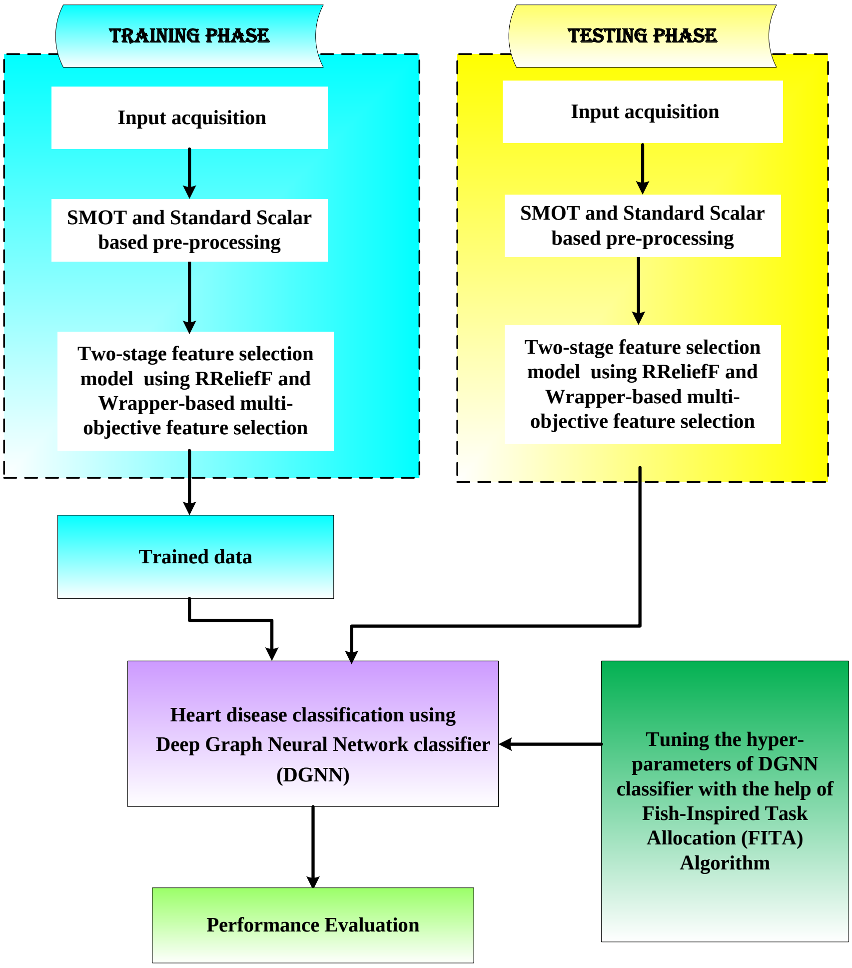

Figure 1 Block diagram of DGNN-FITA-HDD method

However, it takes more time to diagnose and reduced the classification accuracy. For this intention, the Deep Graph Neural Network with Fish-Inspired Task Allocation Algorithm is proposed for identifying the heart diseases accurately with less time.

Initially, the input data samples are pre-processed using SMOT-SS method and the features are selected using two-stage feature selection process. Subsequently, the DGNN approach is used for classifying the absence of heart diseases and presence of heart diseases from the data samples. Finally, the parameters and classification accuracy are enhanced by Fish-Inspired Task Allocation (FITA) algorithm. Figure 1 shows block diagram of DGNN-FITA-HDD technique.

3.1 Pre-processing using Synthetic Minority Oversampling Technique (SMOT) and Standard Scalar (SS) techniques

In this manuscript, SMOT and SS process is utilized for pre-processing input data. Here, the SMOT-SS method reduces the noise and errors in the dataset for improving the data quality based on the procedure of oversampling and class balancing. Also, it retrieves the missing values at the dataset. Initially, the input data is separated into 2 classes: class 0 and class 1 based on mean and standard deviation of data.

The coefficient of each aspect is carried to the equal value by using standard scale to ensure every aspect contains 0 mean and 1 standard deviation. The main objective of over-sampling process is removing the irrelevant data samples for balancing the class of data’s. Moreover, the error or noisy elements in the data is oversampled for developing synthetic data. The size of training position is reduced by diminishing the data classes that are in high number. Finally, the SMOT-SS mechanism is balance the classes by removing unwanted error or noise in the data samples through creating synthetic data of classes for enhancing the data quality. Here, both classes of class 0 and class 1 having equal quantity of data.

3.2 Feature selection using Two-Stage Feature Selection (TSFS) method

The pre-processed data is given to two-stage feature selection technique. The first stage is RReliefF feature selection and 2ndphase is Wrapper-base multiple objective feature selection, which selects features at the data samples.

- RReliefF mechanism

The RReliefF mechanism is based on classic filter for performing feature selection process. The aim of this RReliefF method is to remove the redundant features. Here, the RReliefF method calculates the ability of features for differentiating its neighbor samples based on the degree of variation among the input data and feature. Here, the RReliefF model assign weights to every feature based on the correlation between the dependent variable and data features. The formulation of weights for each feature is calculated using equation (1),

$$F(w) = \left(\frac{P_{dK|dF} P_{dF}}{P_{dK}}\right) - \left(\frac{(1 - P_{dK|dF}) P_{dF}}{1 - P_{dK}}\right) \tag{1}$$

Where F denotes the features, w denotes the weight, dF in P_{dF} represents the different feature values, dK in P_{dK} denotes the different predictions. Thus, the probability of P_{dF} and P_{dK} based nearest instances (NIs) are calculated using equation (2-3),

$$P_{dF} = P(dF \mid NIs) \tag{2}$$

$$P_{dK} = P(dK \mid NIs) \tag{3}$$

Moreover, the condition probability calculation is given in equation (4),

$$P_{dK|dF} = P(dK \mid dFNIs) \tag{4}$$

Here $dFNIs$ represents dF value and its NIs value. The weight of features $F(w)$ are estimated based on the P_{dK}, P_{dK} and $P_{dK|dF}$. Finally, after attaining weight of every feature, feature selection outcome value is regulated using feature weight ranking.

- Wrapper based multiple objective (WMO) feature selection

After completion of RReliefF process, the WMO feature selection is performed in 2nd stage for selecting the input features. Initially, the feature set $(F = F_1, \ldots\ldots\ldots F_N)$ is considered as the input of the Wrapper based multi-objective approach. Subsequently, multi-objective fitness function is calculated and final outcome is obtained based on encoding function of optimal individual, which is given in equation (5),

$$h_k(x) = f(w_k . x + B_k) \tag{5}$$

Where $h_k(x)$ signifies output model in k^{th} node, w_k represents weight matrix of input and output layer, B_k indicates bias vector, $f()$ specifies activation function. Thus, outcomes from equation (5) are considered as input structures to train DGNN method.

3.3 Deep Graph Neural Network (DGNN) for heart disease classification

A DGNN is proposed to classify the heart diseases. Here, the DGNN model analyze the pre-processed data and input features, which classify the data as presence and absence of HD. Normally, DGNN approach has analyzed the interaction characteristics between the samples that is assumed to create the contribution value for user. The better value specifies bigger possibility of detecting HD. Contribution values of every communication record together determine in nature that relations are observed as graph network. Let the overall count of samples used in the dataset as $P_i; i = (1,2,3,\ldots\ldots.n)$ for several patients $U_j; j = (1,2,3,\ldots\ldots.n)$. Subsequently, initial layer $h_{i,j}$ of the DGNN method is calculated using eqn. (6),

$$h_{i,j} = \delta_1 \left\lfloor P_i U_j \left(l_{K1} . X_{i,j} \oplus Y_{i,j}\right) + l_{K2} \right\rfloor \tag{6}$$

Where, l_{K1} and l_{K2} represents input data parameters, ReLU activation function denotes δ_1 and the abstract metrics $X_{i,j}$ and $Y_{i,j}$ are utilized to classify HD and it is calculated below in equation (7-8),

$$X_{i,j} = \delta_1^T .Mp_1\left(V_{i,j}\right) \tag{7}$$

$$Y_{i,j} = \delta_2^T .Mp_2\left(C_{i,j}\right) \tag{8}$$

Where, initial layer is acquired from feature concatenation in the input side $V_{i,j}$ and output side $C_{i,j}$. In this, Mp_1 and Mp_2 are multi-layer used for initializing the features and T represents the transition vector.

Here the processing layers (k) of DGNN is based on totally linked neural mapping function for growing intermediate discrimination vector D_{ij} is given at equation (9),

$$D_{ij} = \delta_1\left\{\frac{1}{|P_i|}\sum_{i=1,j=1}^{n} r_{i,j}.G_{i,j}\left[l_{K1}.h_{ij} + l_{K2}\right]\right\} \tag{9}$$

Where, the classification vector signifies r_{ij}, response function denotes G_{ij} analyses features between samples P_i.

Also, l_{K1} and l_{K2} denotes input sample parameters. Moreover, ReLU activation denotes δ_1. Response function G_{ij} is one then interaction exists. Classification outcomes for each sample is determined by equation (10),

$$y_j = \delta_2\left[D_j.(D_j)^T\right] \tag{10}$$

Here sigmoid activation function denotes δ_2 is determined by equation (11),

$$\delta_2(i,j) = \frac{1}{1+e^{-x}} \tag{11}$$

The range of the outcome of y_j is [0, 1]. Better value of y_j represents error values of data samples that are minimized by DGNN model. To lessen the input imagery size, applying $LK_j^l(b = 1,2,3...t^w)$ mapping function. The input imageries are lessened with the help of 2D filters, also the over fitting issues are lessened through doing this process.

$$LK_j^l = \alpha\left(\sum_{k=1}^{t^{l-1}} LK_j^l * C_{1g}^q + L_k^l\right), b = 1,2,3...t^w \tag{12}$$

Let* denotes convolution operation, L_k^l denotes nonlinear activity function $\theta(.)$. Batch Normalization (BN) is employed to assess hidden neurons on network. It includes 5 channels per module, and changes to 0 padding format. The input size imagery change into 3×3 using max pooling layers. The primary function of DGNN layer is to normalize output of input layer and train the images with each neuron independently. It is given at equation (13),

$$BN_{nor} = \frac{e - nor_{mean}}{e_{Max} - e_{\min}} \tag{13}$$

Where e implies normalize image, nor implies image median, e_{Max} implies maximal value, e_{Min} implies minimal value. This is deemed for rapid training, also offers higher learning outcomes. Max pooling reduces feature mapping function size through down sampling process. The size of the image is lessened to 2×2 as well as [0 0 0 0] padding. It also lessens computational complex. It is given at equation (14),

$$MAX_{\mathrm{Re}s}(i,j) = (L*s)(i,j) = \sum_k\sum_l L(l-q, i-k)\gamma(x,y) \tag{14}$$

Where L the image input size along convolutional model represents i, j, then γ represents training error, $L(l-q, i-k)$ represents loss functioning for transfer learning, t represents convolutions functioning of kernel containing x, y. This process lessens image size on max pooling. The totally linked layers have been utilized for categorize the Heart Disease as presence or absence. It is given at equation (15),

$$Y_{i,j} = \min\left\{\sum_{i,j=1}^{n}\left[-Y_{i,j}\log y_j - \left(1-y_j\right)\log\left(1-y_j\right) + \beta\left\|R_{i,j}\right\|_F^2\right]\right\} \tag{15}$$

Regularization item denotes $\beta\left\|R_{i,j}\right\|_F^2$ for parameters associated to samples. The $Y_{i,j}$ value is given by equation (16),

$$Y_{i,j} = \begin{cases} 1 & ;presence\ \ of\ \ HD \\ 0 & ;absence\ \ of\ \ HD \end{cases} \tag{16}$$

Finally, presence of heart disease and absence of HD in the data samples are classified using eqn. (16). Thus, classification procedure through deep graph neural network has attained with higher outcomes at accuracy, precision, and error rate. Here, these parameters are enhanced by the use of Fish-Inspired Task Allocation (FITA) for improving classification efficiency.

3.4 Fish-Inspired Task Allocation (FITA) Algorithm for optimizing DGNN parameters

The performance measurement of technique is enhanced through DGNN parameters using FITA algorithm. Here, the FITA algorithm aims to raise the accuracy for enhancing the performance of heart disease classification. The Fish-Inspired Task Allocation (FITA) Algorithm is based on the behavior of biologic fish. Here, the search space is constrained and every possible location denotes a promising solution for the problem. This FITA Algorithm can enhance the performance using the fitness function, swimming, feeding and breeding operators. The achievement of fish during search process is specified based on the weight, so promising areas are inferred from the regions where larger groups of fish are placed. Moreover, the FITA algorithm is performed based on swimming, feeding and breeding operators. Figure 2 depicts the flow chart representation of FITA process. The entire FITA process is described at following steps.

Step 1: Initialization process

The initial input parameters are set depending on the weights and locations of every fish. The decision variables and iterations are then calculated to improve categorizations procedure parameters and address complexity difficulties. Thus, initialization process in FITA of input parameters are mentioned in equation (17),

$$P = \mathrm{int}(X_{i,j}, Y_{i,j}, D_{i,j}) \tag{17}$$

Where the input image data samples for every patient are denoted as $i, j = (1,2......n)$ and $X_{i,j}, Y_{i,j}, D_{i,j}$ are denotes the input, output, and discrimination vector parameters.

Step 2: Random generation

In this step, the location for all fish is randomly generated and set the weights for each individual as one. Also, the swimming operation of FITA algorithm calculates the random movement of every individual fish. The proposed FITA algorithm analyses the features of every data samples in this step.

Step 3: Fitness function to optimize DGNN parameters

The FITA fitness function is computed to improve DGNN parameters for attaining better classification results. Fitness function (F) is determined by equation (18),

$$F_{i,j} = f\left(P.x_{i,j}(t+1) - f[x_{i,j}(t)]\right) \tag{18}$$

Where, $[x_{i,j}(t)]$ denotes every fish position, $x_{i,j}(t+1)$ denotes the updated position and $f(P)$ represents the fitness of initialized parameters. Here, the updated position of each fish is used for analyzing the features of each data samples. If the possible outcome is not attained using fitness function, then the feeding operator is initialized.

Step 4: Feeding operator for calculating variation of fish weights

The feeding operator defines the variation of fish weight. Normally, every fish is born with equal weight and the weight is started with one. Here, the weight of every fish is increased or decreased based on success/failure of fish movement. This procedure is applied to raise the performance and decrease the error function during classification. The variation of weights $w_{i,j}(t+1)$ for each fish is calculated using eqn.(19),

$$w_{i,j}(t+1) = w_{i,j}(t) + \frac{\Delta F_{i,j}}{\max(\Delta F_{i,j})} \tag{19}$$

here $w_{i,j}(t)$ denotes weight of each individuals, and $\Delta F_{i,j}$ is represents the fitness difference.

Step 5: Swimming operator for updating position of each individuals

Swimming behavior is associated to every main individuals and collective behaviors. Here, swimming behavior is considered as the comprehensive form of survival-related reaction. It aims to simulate the cohesive and expressive collective movement created by all the fish in the swarm. Moreover, the position of each individuals are updated using eqn (20),

$$x_{i,j}(t+1) = x_{i,j}(t) + m(t) \tag{20}$$

Where $x_{i,j}(t)$ is the actual position of each individuals, $x_{i,j}(t+1)$ specifies the updated position of each individuals and $m(t)$ represents the resulting direction that is calculated using eqn (21).

$$m(t) = \frac{\sum_{i,j=0}^{n} \Delta x_{i,j}(t+1)\Delta F_{i,j}}{\sum_{i,j=1}^{n} \Delta F_{i,j}} \quad (21)$$

Here $\Delta x_{i,j}(t+1)$ denotes variation of fish position by movements, and $\Delta F_{i,j}$ is represents the fitness difference.

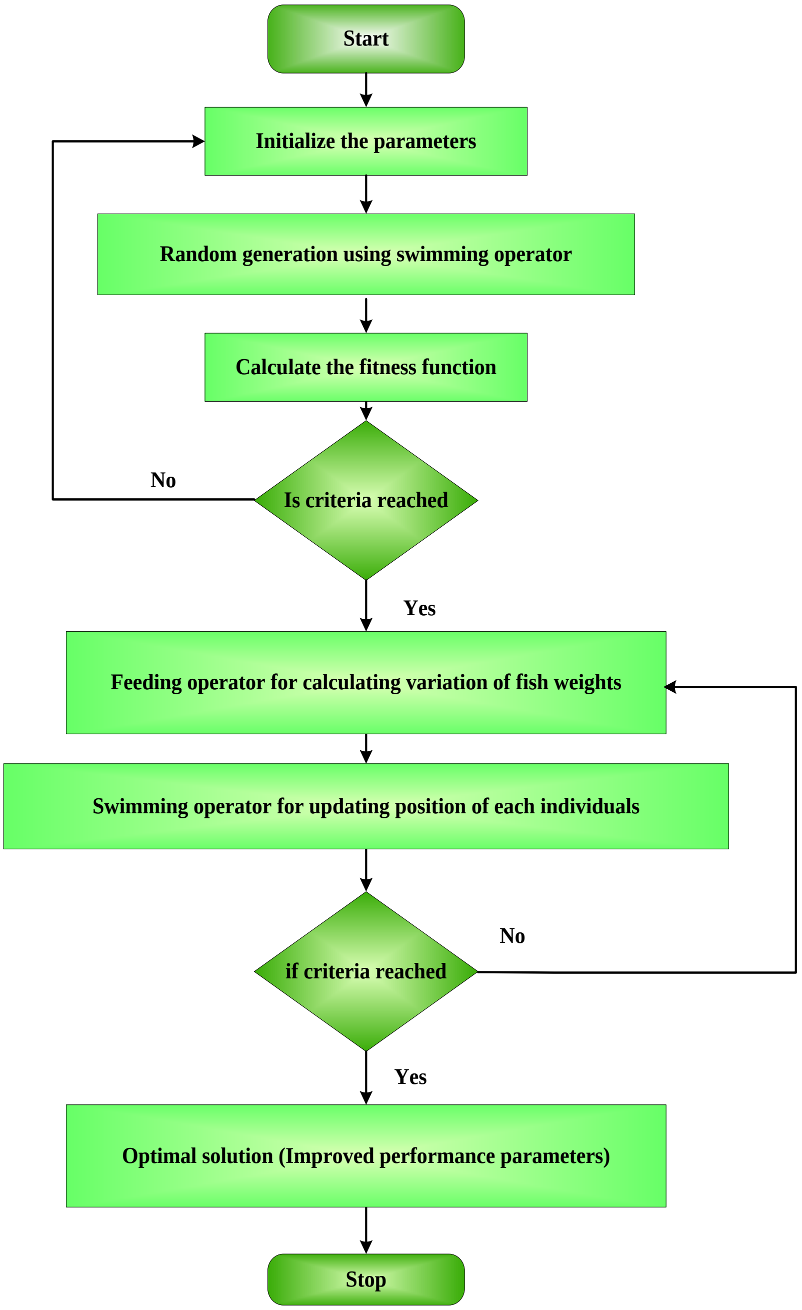

Figure 2 Flow chart representation of FITA

Step 6: Termination

The optimal solution is attained from eqn (18) that enhance the performance metrics of the DGNN classification. Therefore, the heart disease classification accuracy has improved by proposed DGNN- FITA mechanism.

4. Result and discussion

The performance of DGNN-FITA-HDD method is described. The proposed technique successfully categorizes cardiac illnesses based on given data. In MATLAB, this methodology's simulation is performed. The DGNN-FITA-HDD method has achieved higher results under performance metrics, like accuracy, f-measure, error rate, precision, specificity sensitivity. Table 1 shows the simulation parameter of proposed technique.

4.1 Dataset description

Cleveland database from machine learning repository center at UCI (https://archive.ics.uci.edu/ml/datasets/heart+disease) is utilized for classifying presence of HD through DGNN-FITA-HDD method. The overall count of instances at the dataset is 303 with 75 attributes that involves patients identification number, age, sex. In this, 50% of input data is employed for testing, 50% of remaining data samples are employed for training.

In the proposed DGNN-FITA-HDD method, the input data samples are utilized to classify HD and the balanced data distribution is depicted in Figure 3. Figure 4 depicts the imbalanced data distribution.

Table 1 Simulation parameter

Parameter	Value
Total count of data samples	500
Maximum number of sample in each class	400
Iteration	500
Iterations per epoch	140
Elapsed time	10 minutes 45 seconds
Hardware resource	Single CPU
Learning rate	0.001

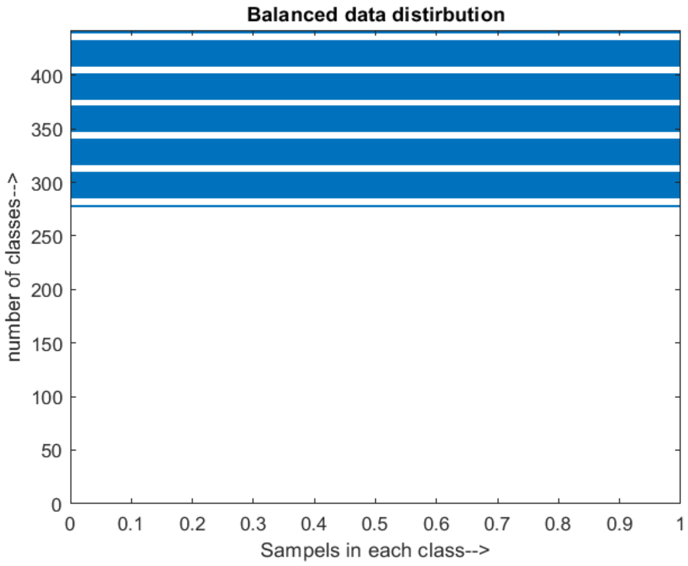

Figure 3 Balanced data distribution

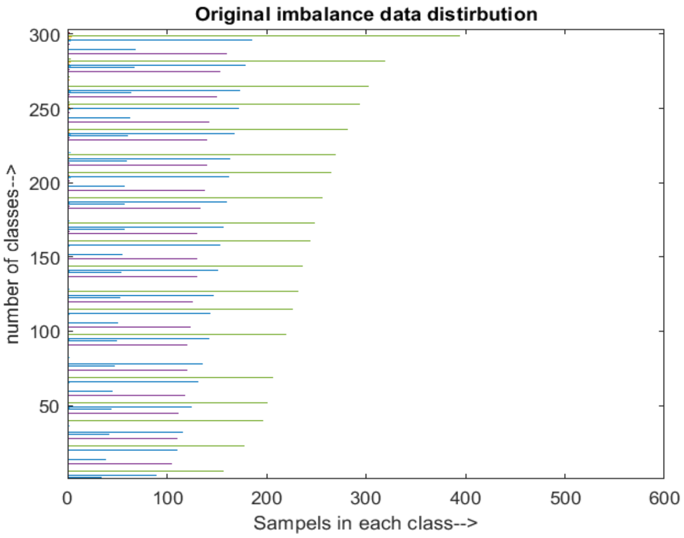

Figure 4 Imbalanced data distribution

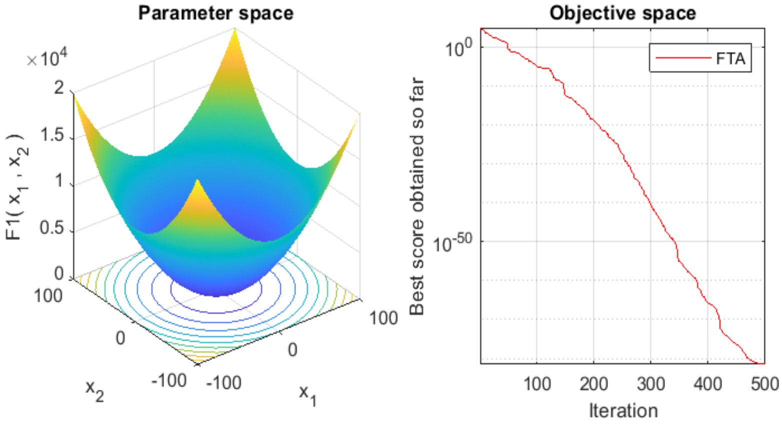

Figure 5 Parameter space and objective space curve for FITA benchmarks function

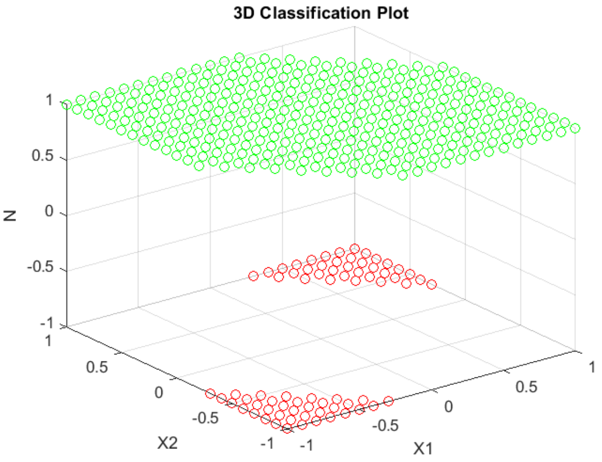

Figure 6 3D Classification plot using proposed DGNN- FITA method

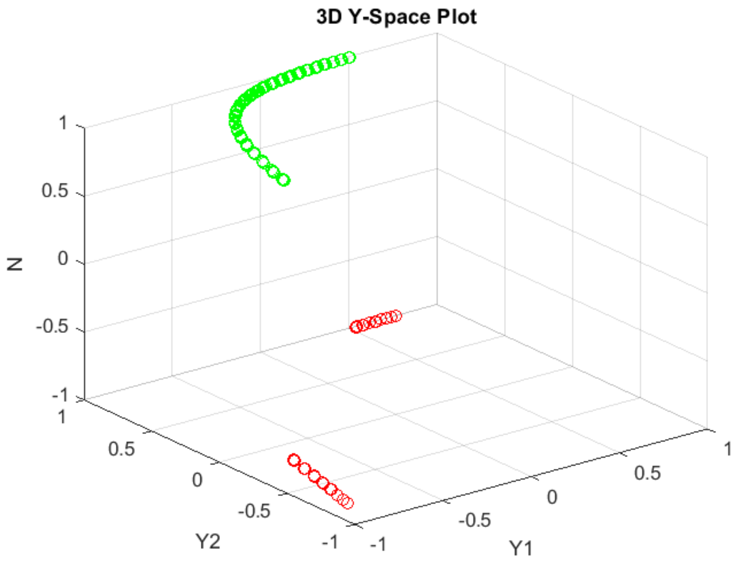

Figure 7 3D Y-space Classification plot using proposed DGNN- FITA method

In Figure 5, the objective space curve is among the attained best fitness value and iterations. The curve depicts that the objective function minimal value based on the optimization.

Also, the classification plot in 3D representation using proposed DGNN- FITA method is shown in Figure 6.

Additionally, the classification plot in 3D Y-space representation using proposed DGNN-FITA method is shown in Figure 7.

4.2 Performance metrics

The performance metrics are examined for checking efficacy of proposed method. To scale performance metrics, the confusion matrix is deemed.

- Accurate HD categorizations from data samples denotes true positive (T_P) .
- Inaccurate HD samples prediction into healthy sample represents false positive (F_P)
- Accurate healthy samples prediction from data samples denotes true negative value (T_N) .
- Inaccurate healthy samples prediction into HD sample represents false negative (F_N) .

4.2.1 Calculation of Accuracy

Accuracy is determined to detect the efficacy of DGNN- FITA method when categorizing heart diseases from data samples, which is calculated using eqn (22),

$$Acc(A)=\frac{T_P+T_N}{T_P+F_P+T_N+F_N} \tag{22}$$

4.2.2 Calculation of Precision

Precision calculation is used for measuring efficacy of DGNN- FITA technique. It is given by equation (23),

$$\Pr=\frac{T_P}{T_P+F_P} \tag{23}$$

4.2.3 Calculation of Sensitivity

Sensitivity is used to identify the presence of heart disease from the data samples using equation (24),

$$Se=\frac{T_P}{T_P+F_N} \tag{24}$$

4.2.4 Calculation of Specificity

Specificity computation is employed for detecting the absence of heart disease from the data samples, which is calculated using eqn (25),

$$Sp=\frac{T_N}{T_N+F_P} \tag{25}$$

4.2.5 Calculation of F-score

Accuracy value and recall must be balanced throughout f-score computation. It is given by eqn (26),

$$F-score=\frac{\mathrm{Re}\,.\,\mathrm{Pr}}{\mathrm{Re}+\mathrm{Pr}} \tag{26}$$

4.2.6 Calculation of error rate

Error rate computation is determined by inaccurate prediction and HD categorization through proposed approach. It is given at equation (27),

$$Error\_rate=\frac{F_P+F_N}{Total} \tag{27}$$

4.3 Comparative analysis of performance metrics

Figure 8-13 displays simulation outcomes of deep graph neural network with fish-inspired task allocation approach is proposed for heart disease diagnosis. Performance metrics are comparing to the existing higher performance in heart disease categorization utilizing advanced supercomputing method along cluster based enhanced deep genetic approach (EDGA-AHHO-HDD) (Bakhsh, 2021), artificial plant optimization approach to identify heart rate & heart disease utilizing machine learning (XGB-MAPO-Hbnm,./DD) (Sharma et al., 2020), Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis (AGAFL-HDD) (Reddy et al., 2020) and efficient prediction of cardiovascular disease utilizing machine learning approaches along relief and LASSO methods (RFBM-HDD (Ghosh et al., 2021) techniques respectively.

In Figure 8, the performance of proposed DGNN-FITA-HDD -HDD method is compared with existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. The proposed DGNN-FITA-HDD -HDD method without feature selection process attains 58.06%, 42.26%, 63.33%and 8.64% high accuracy than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. Moreover, the proposed DGNN-FITA-HDD –HDD method with feature selection process attains 11.49%, 24.5%, 49.23% and 13.41% high accuracy than the existing EDGA-AHHO and XGB-MAPO methods. From the outcome, the proposed method with feature selection process has achieved higher accuracy rate for HD classification.

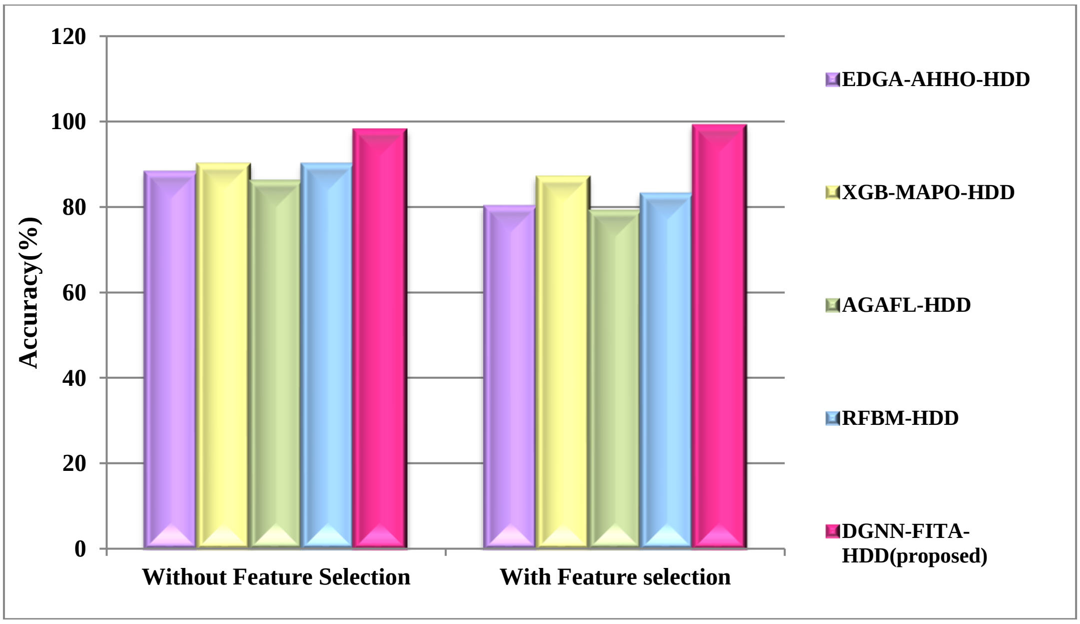

Figure 8 Comparison of accuracy

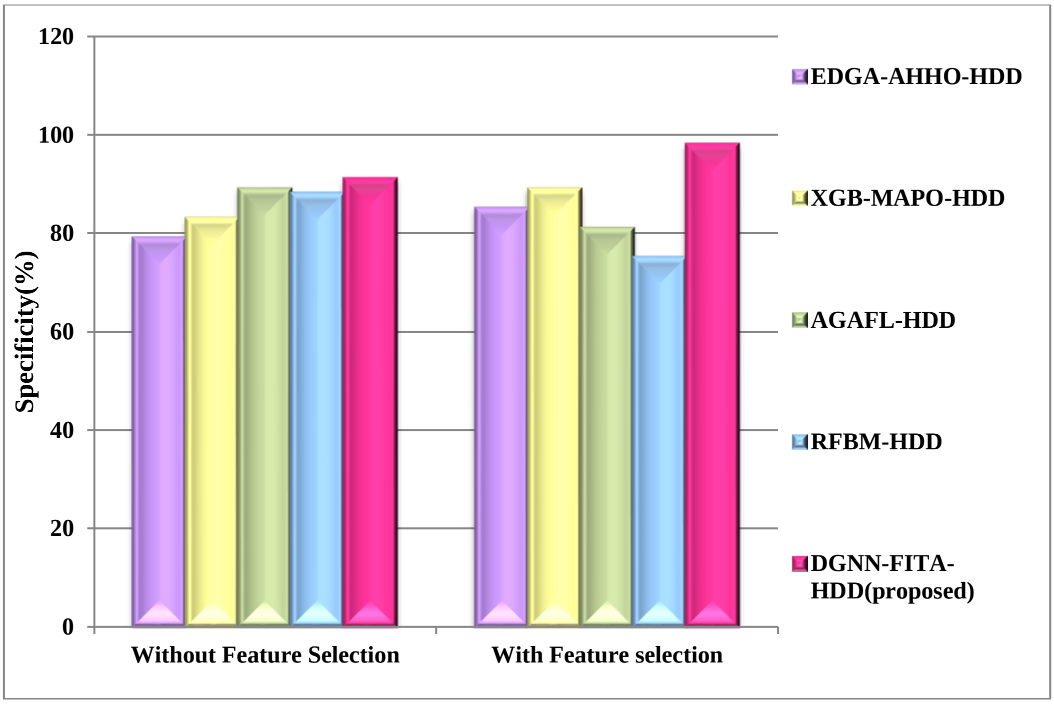

Figure 9 Comparison of specificity

In Figure 9, displays performance of DGNN-FITA-HDD technique is comparing to the existing EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD methods. The proposed DGNN-FITA-HDD method without feature selection process attains 16.21%, 22.33%, 43.16%and 48.63% high specificity than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. Moreover, the proposed DGNN-FITA-HDD method with feature selection process attains 21.31%, 14.06%, 39.16% and 21.36% high specificity than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. From the outcome, the proposed method with feature selection process has achieved higher specificity rate for HD classification.

In Figure 10, the performance of proposed DGNN-FITA-HDD -HDD method is compared with existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. The proposed DGNN-FITA-HDD-HDD method without feature selection process attains 23.22%, 61.27%, 58.06%, and 22.5% high sensitivity than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. Moreover, the proposed DGNN-FITA-HDD –HDD method with feature selection process attains 41.26%, 37.77%, 51.43%, and 75.36% high sensitivity than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. From the outcome, the proposed method with feature selection process has achieved higher accuracy rate for HD classification.

In Figure 11, the performance of proposed DGNN-FITA-HDD -HDD method is compared with existing methods, like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. The proposed DGNN-FITA-HDD -HDD method without feature selection process attains 16.31%, 14.913%, 16.056% and 15.812% high precision than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. Moreover, the proposed DGNN-FITA-HDD -HDD method with feature selection process attains 11.31%, 15.988%, 14.542% and 12.478% high precision than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. From the outcome, the proposed method with feature selection process has achieved higher precision rate for HD classification.

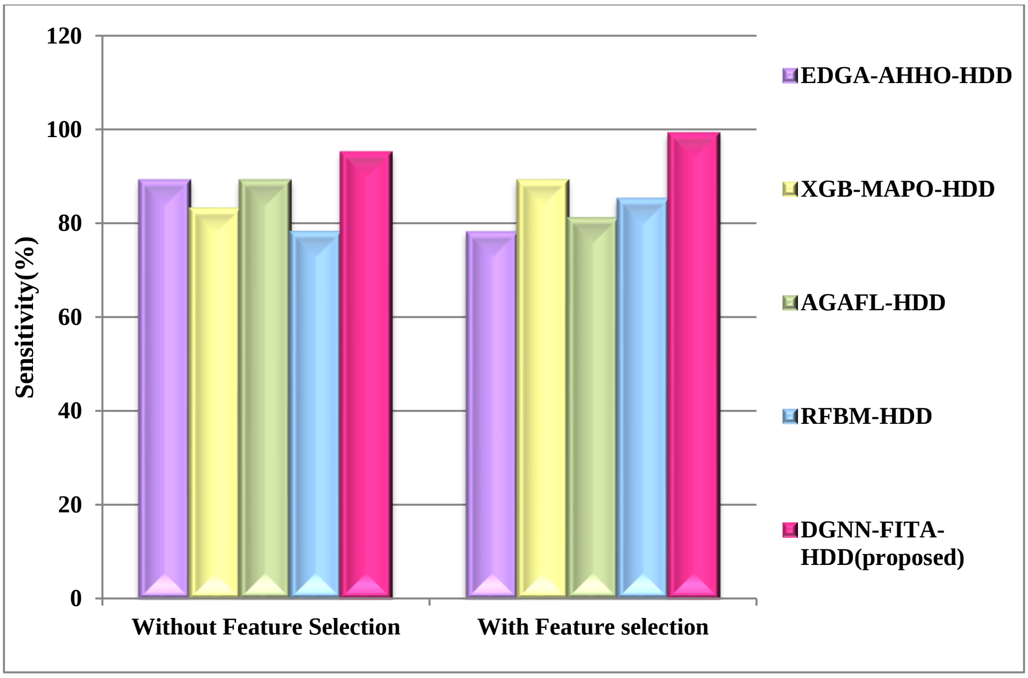

Figure 10 Comparison of sensitivity

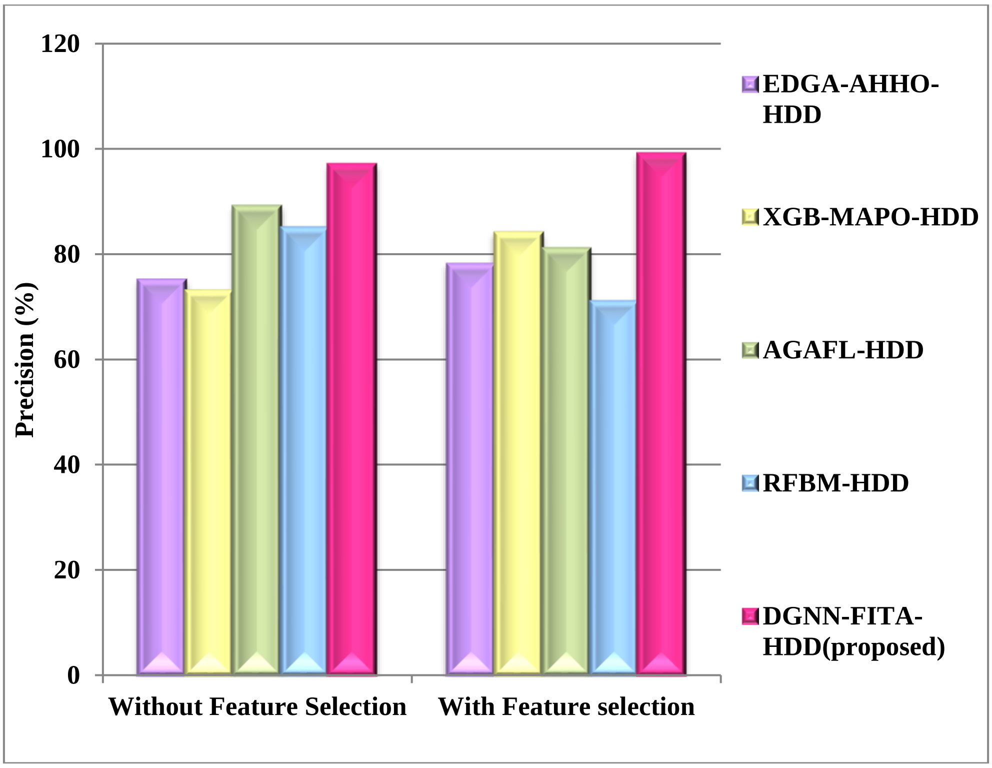

Figure 11 Comparison of precision

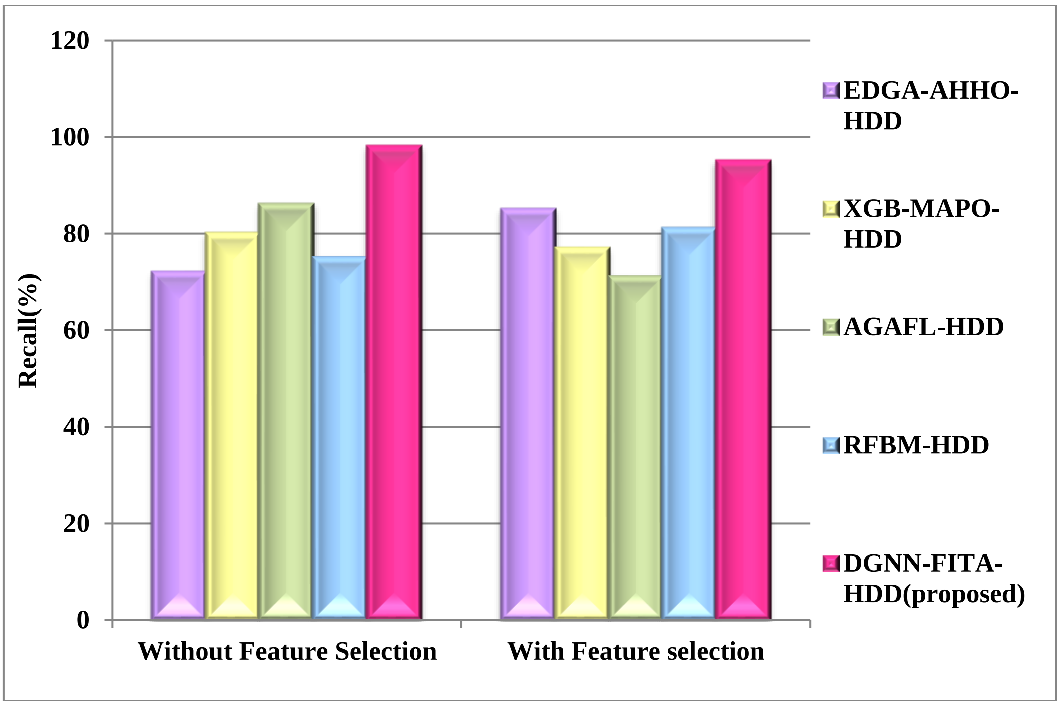

Figure 12 Comparison of Recall

In Figure 12, the performance of proposed DGNN-FITA-HDD -HDD method is compared with existing methods, like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. The proposed DGNN-FITA-HDD -HDD method without feature selection process attains 31.27%, 48.22%, 65.23% and 14.84% high recall than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. Moreover, the proposed DGNN-FITA-HDD -HDD method with feature selection process attains 19.32%, 14.26%, 14.407% and 20.391% high recall than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. From the outcome, the proposed method with feature selection process has achieved higher recall rate for HD classification.

In Figure 13, displays performance of DGNN-FITA-HDD -HDD technique is comparing with existing EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. The proposed DGNN-FITA-HDD -HDD method without feature selection process attains 29.41%, 16.3%, 11.327% and 15.209% high F-Measure than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD methods. Moreover, the proposed DGNN-FITA-HDD -HDD method with feature selection process attains 29.11%, 14.97%, 14.782% and 37.46% high F-Measure than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. From the outcome, the proposed method with feature selection process has achieved higher F-Measure rate for HD classification.

In Figure 14, the performance of proposed DGNN-FITA-HDD -HDD method is compared with existing EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD methods. The proposed DGNN-FITA-HDD -HDD method without feature selection process attains 15.98% 9.316%, 11.785% and 22.13% low error rate than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. Moreover, the proposed DGNN-FITA-HDD -HDD method with feature selection process attains 48.22%, 16.021%, 19.84% and 51.6% low error rate than the existing methods like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively. From the outcome, the proposed method with feature selection process has achieved lower error rate for HD classification.

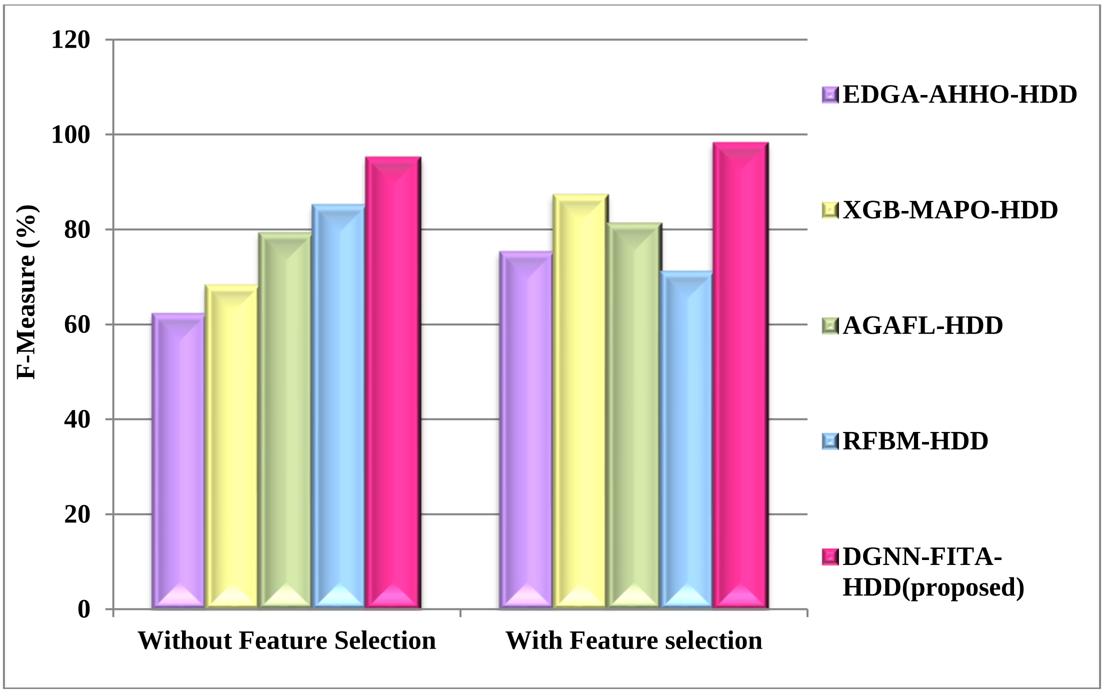

Figure 13 Comparison of F-Measure

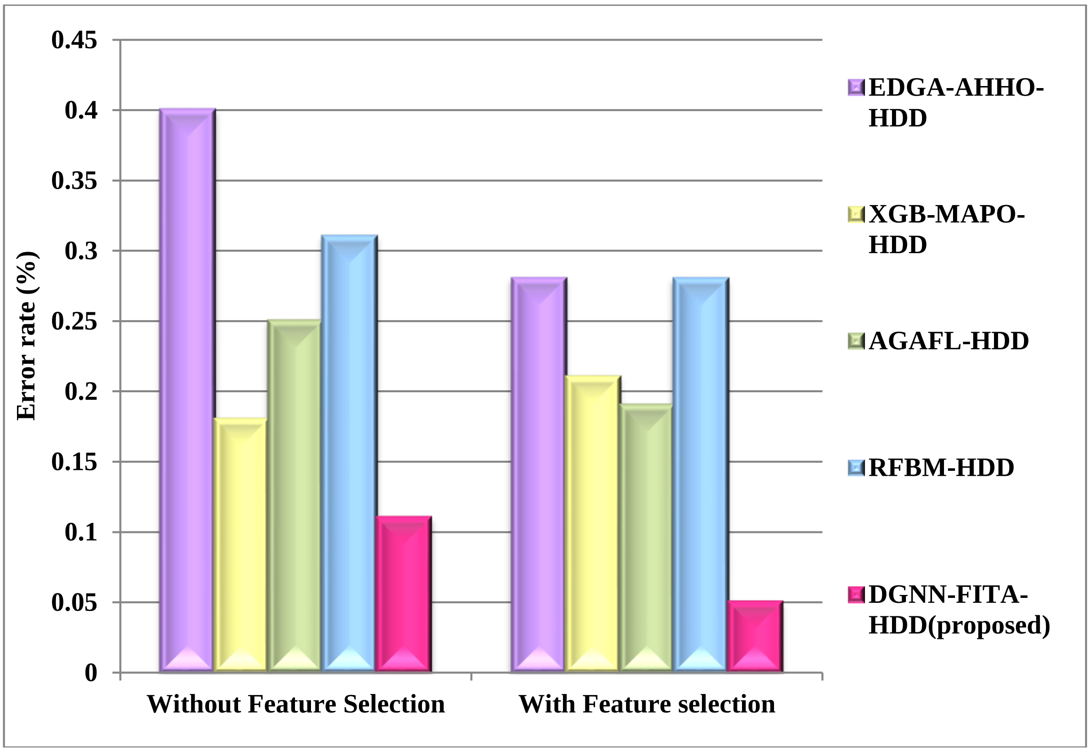

Figure 14 Comparison of Error rate

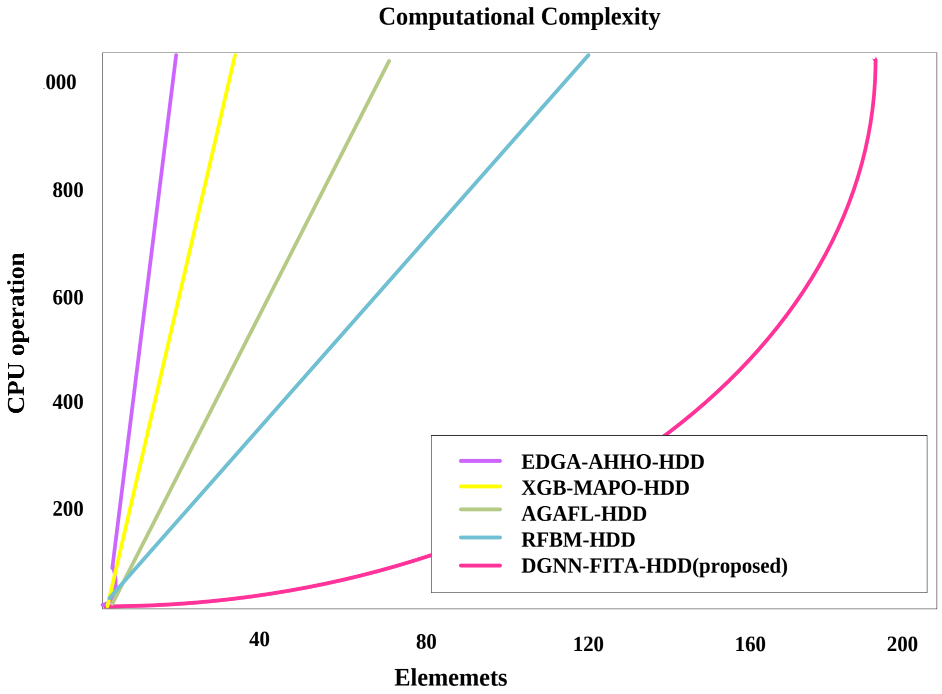

Figure 15 Computational Complexity

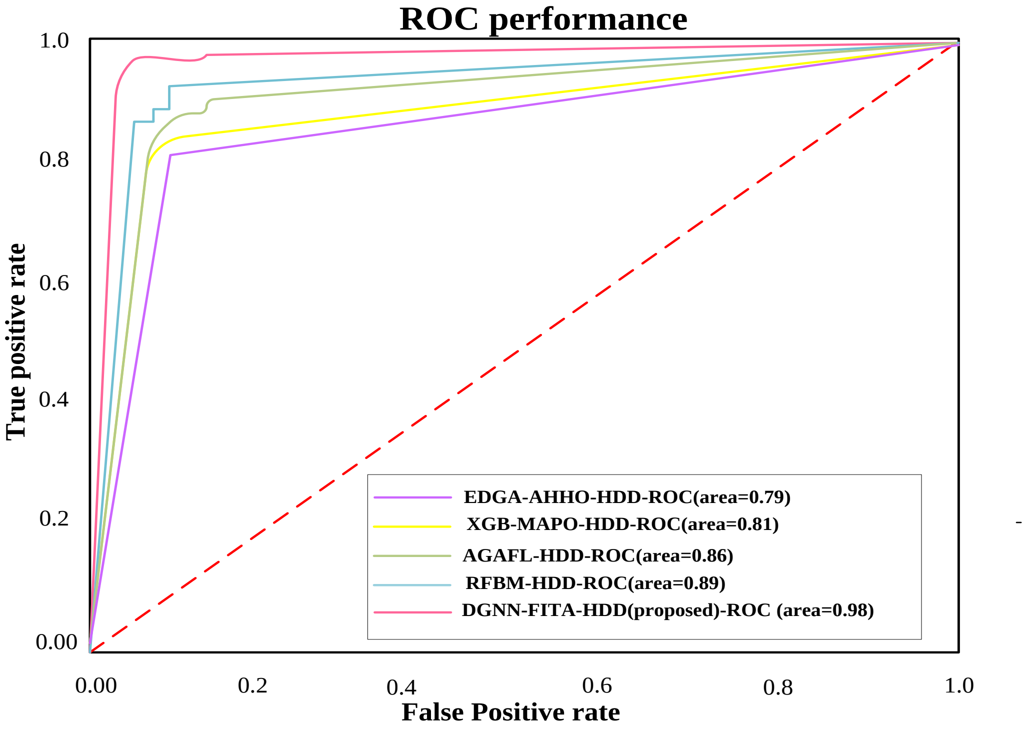

Figure 16 ROC curve for Heart Disease Diagnosis

Figure 15 shows computational complexity of DGNN-FITA-HDD -HDD CPU operation time and memory usage are linearly upgraded based on input data value. The proposed DGNN-FITA-HDD -HDD has less computational complex compared to the existing EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD methods respectively.

Figure 16 displays ROC curve for Heart Disease Diagnosis. The ROC of DGNN-FITA-HDD -HDD technique offers 2.89%, 6.59%, 8.56%, 6.34% better AUC than existing EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD techniques respectively.

5. Conclusion

A Deep Graph Neural Network with Fish-Inspired Task Allocation (DGNN-FITA-HDD) Algorithm is successfully implemented in MATLAB for classifying the heart disease. Two-Stage Feature Selection technique selects the most important features from pre-processing output. Also, Fish-Inspired Task Allocation mechanism enhanced the classification parameters of DGNN. The proposed DGNN-FITA-HDD -HDD approach has attained better performance under accuracy, specificity, sensitivity, f-measure, and error rate. Here, 32.11%, 43.29, 15.57% and 26.31% improved specificity, 26.55%, 37.74%, 41.32% and 12.56% high sensitivity, 53.26%, 32.03%, 16.29% and 36.52% high precision and 26.52%, 11.55%, 22.18% and 31.26% high F-measure are attained by the proposed DGNN-FITA-HDD -HDD approach with feature selection method when compared to the existing methods, like EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD. Moreover, 13.96%, 27.56%, 33.26% and 47.01% high specificity, 11.4%, 29.6%, 36.12% and 19.26% high sensitivity, 13.4%, 52.6%, 28.36% and 12.06% high precision, and 15.76%, 20.36%, 36.03% and 15.29% high F-measure are attained by the proposed DGNN-FITA-HDD approach without feature selection method compared to the existing EDGA-AHHO-HDD, XGB-MAPO-HDD, AGAFL-HDD and RFBM-HDD respectively.

In the field of heart disease diagnosis, there are many potential directions for future research. One promising area is the use of artificial intelligence and machine learning algorithms to improve accuracy and speed of diagnosis. Biomarkers and genetic testing may also be able to improve the accuracy of diagnosis, while wearable technology can provide doctors with more data to monitor patients for early signs of heart disease. As our understanding of the genetic and environmental factors that contribute to heart disease improves, there is a growing interest in developing personalized treatments based on individual patients' unique risk factors and genetic profiles. Finally, telemedicine can improve access to heart disease diagnosis and treatment, particularly for patients in remote or underserved areas, ultimately improving patient outcomes and quality of life. Overall, these various avenues of research offer exciting possibilities for improving heart disease diagnosis and treatment at future.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

6. References

Abdel-Basset, M., Gamal, A., Manogaran, G., Son, L. H., & Long, H. V. (2020). A novel group decision making model based on neutrosophic sets for heart disease diagnosis. *Multimedia Tools and Applications*, *79*(15), 9977-10002. https://doi.org/10.1007/s11042-019-07742-7

Alhaqbani, A., Kurdi, H., &Youcef-Toumi, K. (2020). Fish-inspired task allocation algorithm for multiple unmanned aerial vehicles in search and rescue missions. *Remote Sensing*, *13*(1), Article 27. https://doi.org/10.3390/rs13010027

Bakhsh, A. A. (2021). High-performance in classification of heart disease using advanced supercomputing technique with cluster-based enhanced deep genetic algorithm. *The Journal of Supercomputing*, *77*(9), 10540-10561. https://doi.org/10.1007/s11227-021-03689-5

Einstein, A. J., Shaw, L. J., Hirschfeld, C., Williams, M. C., Villines, T. C., Better, N., ... & INCAPS COVID Investigators Group. (2021). International impact of COVID-19 on the diagnosis of heart disease. *Journal of the American College of Cardiology*, *77*(2), 173-185. https://doi.org/10.1016/j.jacc.2020.10.054

Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, *9*, 19304-19326. https://doi.org/10.1109/ACCESS.2021.3053759

Gokulnath, C. B., &Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, *22*(6), 14777-14787. https://doi.org/10.1007/s10586-018-2416-4

Guo, Z., Tang, L., Guo, T., Yu, K., Alazab, M., &Shalaginov, A. (2021). Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace. *Future generation computer systems*, *117*, 205-218.https://doi.org/10.1016/j.future.2020.11.028

Khan, M. A. (2020). An IoT framework for heart disease prediction based on MDCNN classifier. *IEEE Access*, *8*, 34717-34727. https://doi.org/10.1109/ACCESS.2020.2974687

Khan, M. A., &Algarni, F. (2020). A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. *IEEE Access*, *8*, 122259-122269. https://doi.org/10.1109/ACCESS.2020.3006424

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., &Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*, *8*, 107562-107582. https://doi.org/10.1109/ACCESS.2020.3001149

Liu, T., Tian, Y., Zhao, S., Huang, X., & Wang, Q. (2020). Residual convolutional neural network for cardiac image segmentation and heart disease diagnosis. *IEEE Access*, *8*, 82153-82161. https://doi.org/10.1109/ACCESS.2020.2991424

Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and Internet of Things enabled disease diagnosis model for smart healthcare systems. *IEEE Access*, *9*, 45137-45146. https://doi.org/10.1109/ACCESS.2021.3066365

Nilashi, M., Ahmadi, H., Manaf, A. A., Rashid, T. A., Samad, S., Shahmoradi, L., ... & Akbari, E. (2020). Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. *International Journal of Fuzzy Systems*, *22*(4), 1376-1388. https://doi.org/10.1007/s40815-020-00828-7

Niu, T., Wang, J., Lu, H., Yang, W., & Du, P. (2020). Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications*, *148*, Article 113237. https://doi.org/10.1016/j.eswa.2020.113237

Nourmohammadi-Khiarak, J., Feizi-Derakhshi, M. R., Behrouzi, K., Mazaheri, S., Zamani-Harghalani, Y., &Tayebi, R. M. (2020). New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection. *Health and Technology*, *10*, 667-678. https://doi.org/10.1007/s12553-019-00396-3

Rajesh, P., Muthubalaji, S., Srinivasan, S., & Shajin, F. H. (2022). Leveraging a Dynamic Differential Annealed Optimization and Recalling Enhanced Recurrent Neural Network for Maximum Power Point Tracking in Wind Energy Conversion System. *Technology and Economics of Smart Grids and Sustainable Energy*, *7*(1), 1-15. https://doi.org/10.1007/s40866-022-00144-z

Rajesh, P., Shajin, F. H., Mouli Chandra, B., &Kommula, B. N. (2021). Diminishing Energy Consumption Cost and Optimal Energy Management of Photovoltaic Aided Electric Vehicle (PV-EV) By GFO-VITG Approach. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 1-19. https://doi.org/10.1080/15567036.2021.1986606

Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, *7*(3), 263-275. https://doi.org/10.1007/s40860-021-00133-6

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Rajput, D. S., Kaluri, R., & Srivastava, G. (2020). Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evolutionary Intelligence*, *13*, 185-196. https://doi.org/10.1007/s12065-019-00327-1

Saqlain, S. M., Sher, M., Shah, F. A., Khan, I., Ashraf, M. U., Awais, M., & Ghani, A. (2019). Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. *Knowledge and Information Systems*, *58*(1), 139-167. https://doi.org/10.1007/s10115-018-1185-y

Shaji, S. P. (2019, April). Predictionand Diagnosis of Heart Disease Patients using Data Mining Technique. In *2019 international conference on communication and signal processing (ICCSP)* (pp. 0848-0852). IEEE.

Shajin, F. H., & Rajesh, P. (2022). FPGA Realization of a Reversible Data Hiding Scheme for 5G MIMO-OFDM System by Chaotic Key Generation-Based Paillier Cryptography Along with LDPC and Its Side Channel Estimation Using Machine Learning Technique. *Journal of Circuits, Systems and Computers*, *31*(05), 2250093. https://doi.org/10.1142/S0218126622500931

Shajin, F. H., Rajesh, P., & Raja, M. R. (2021). An Efficient VLSI Architecture for Fast Motion Estimation Exploiting Zero Motion Prejudgment Technique and a New Quadrant-Based Search Algorithm in HEVC. *Circuits, Systems, and Signal Processing*, *41*, 1751–177. https://doi.org/10.1007/s00034-021-01850-2

Sharma, P., Choudhary, K., Gupta, K., Chawla, R., Gupta, D., & Sharma, A. (2020). Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning. *Artificial intelligence in medicine*, *102*, Article 101752. https://doi.org/10.1016/j.artmed.2019.101752

Tesson, S., Butow, P. N., Marshall, K., Fonagy, P., & Kasparian, N. A. (2022). Parent-child bonding and attachment during pregnancy and early childhood following congenital heart disease diagnosis. *Health Psychology Review*, *16*(3), 378-411. https://doi.org/10.1080/17437199.2021.1927136

ThangaSelvi, R., &Muthulakshmi, I. (2021). An optimal artificial neural network based big data application for heart disease diagnosis and classification model. *Journal of Ambient Intelligence and Humanized Computing*, *12*(6), 6129-6139. https://doi.org/10.1007/s12652-020-02181-x

Wang, J., You, T., Yi, K., Gong, Y., Xie, Q., Qu, F., ... & He, Z. (2020). Intelligent diagnosis of heart murmurs in children with congenital heart disease. *Journal of healthcare engineering*, Article 9640821. https://doi.org/10.1155/2020/9640821