

Review of Computer Engineering Research

2023 Vol. 10, No. 4, pp. 150-164

ISSN(e): 2410-9142


ISSN(p): 2412-4281

DOI: 10.18488/76.v10i4.3549

© 2023 Conscientia Beam. All Rights Reserved.



Whale optimization approach for early heart disease prediction based on FCM using DCNN

 **Rajakumar Raja**
Aswathi^{1*}
Kolappanpillai Pazhani
Kumar²
Bagavathiperumal
Ramakrishnan³

^{1,2,3}Department of Computer Science, S. T. Hindu College, Nagercoil, India.

¹Email: rrajaaswathi@gmail.com

²Email: skpk73@gmail.com

³Email: ramsthc@gmail.com



(+ Corresponding author)

ABSTRACT

Article History

Received: 4 July 2023

Revised: 27 September 2023

Accepted: 19 October 2023

Published: 13 December 2023

Keywords

Data mining

DCNN

FCM

Machine learning

WOA.

Heart disease is the leading cause of death worldwide. It has an impact on not only the health of patients but also the economies and expenses of the countries. Numerous machine learning and data mining approaches are being developed and explored currently in order to predict various diseases. This paper aims to address the pressing global issue of heart disease by leveraging machine learning and data mining techniques. Specifically, it focuses on utilizing a Fuzzy C means (FCM) approach for attribute segmentation, employing the Whale Optimization Algorithm (WOA) for feature selection, and utilizing Deep Convolutional Neural Networks (DCNNs) for medical diagnosis and early prediction. In this study, the initial stage involves segmenting patient records' attributes using the FCM method. Subsequently, high-ranking features are selected through the WOA algorithm. These segmented features are then input into DCNNs to construct a robust medical diagnosis system and enable early-stage prediction. The DCNNs autonomously extract crucial features without human intervention, enhancing the accuracy of disease prediction. The performance evaluation of the proposed classifier is conducted using the Python platform, with the DCNN achieving an impressive accuracy level of 90.12% during testing. This indicates the DCNN's capability to accurately predict the presence or absence of cardiac disease, showcasing its potential as an effective tool in healthcare. The integration of FCM attribute segmentation, WOA feature selection, and DCNN-based prediction holds significant practical implications. It offers healthcare professionals a valuable tool for diagnosing and predicting heart disease early, potentially saving lives.

Contribution/Originality: This study stands out for its innovative integration of the FCM approach for segmentation, WOA for feature selection, and DCNNs for prediction. This approach's originality lies in its potential to enhance accuracy in predicting heart disease while reducing manual interpretation, setting it apart from previous works.

1. INTRODUCTION

Currently, heart disease is one of the most widespread illnesses in the world. It is predicted that 17.9 million deaths, or 15% of all natural deaths, occurred as a result of heart disease in 2017. By monitoring the levels of different health indicators such as cholesterol, glucose level, blood pressure, and heart rate, early stages of heart disease can be identified [1-3]. Cardiovascular disease affects not only the health of people but also the economies and costs of nations. Therefore, the diagnosis and treatment of heart illness are always crucial. An initial prognosis

can help a patient obtain more effective and proper treatment. Therefore, the demand to create such an early diagnosis and prediction system for medicine grows every day. One of this system's most crucial features should be great accuracy outputs at a low running cost [4, 5].

Medical professionals' activity produces a variety of datasets that can be examined to identify the main characteristics for detecting heart attacks. Unfortunately, these datasets aren't being used to their full potential right now [6]. The main objective of the research is to make use of those real-life datasets in a way that could aid in the timely prediction of a probable heart attack. Here, the data analysis and data mining approaches are available. To explore hidden patterns in medical data sets and estimate the patient's future condition, data mining has made it possible to combine classification approaches and give computerized training on the dataset [7]. Thus, through the analysis of medical data, it is feasible to provide perceptions into a patient's background and offer clinical assistance. So, the data mining methods are deployed to establish the heart disease prediction exactly [7]. Several studies have been proposed to forecast heart disease in the early stages, which are listed below.

An excellent framework for incident heart failure diagnosis utilizing routine Electronic Health Record (EHR) is described by the author in Rao, et al. [8], opening up an interesting emerging area of study for the forecasting of additional complex conditions. In-depth phenotype populations ought to be examined with the framework and evaluation necessary to find novel illness processes and trends in other difficult situations. However, as with conventional statistical models, extreme caution ought to be exercised when making assumptions about causation. The heart disease risk forecast technique using Deep Neural Network (DNN) models is presented in ref [9] and used with the KNHANES samples. With the help of the Principle Component Analysis (PCA) and Variational Auto Encoders (VAE) frameworks, the proposed approach solved the issue of creating effective trained data by identifying and enlarging significantly influenced subsets that impair effectiveness of the models. The performance is greatly enhanced by the DNN learned in the PCA-partitioned learning groups. The method suggested has the drawback of not allowing missing values. Determining the ideal threshold in this module is also difficult. CardioXNet, a unique Convolution Recurrent Neural Network (CRNN) network, has been suggested in ref [10] for the programmed diagnosis of various diseases regarding heart classes without undergoing any Phonocardiogram (PCG) signal preprocessing. With an accuracy score of 99.6%, the framework displayed state-of-the-art performance, beating all earlier research on provided PCG data by a significant margin. But while the current system performs satisfactorily on PCG datasets that mimic real-life settings, the work can still be enhanced if more substantial PCG datasets with a variety of heart disease observations are made available.

A new method for Chronic Heart Failure (CHF) recognition from PCG audio records has been described by the author in [11]. The technique integrates Deep Learning (DL) with conventional Machine Learning (ML). DL trains from both spectral representations and the time domain of the signal, whereas conventional ML trains from a huge amount of defined features. As long as domain-specific labelled data are available, the recommended approach proves to be reliable and beneficial for identifying other heart-sound categorization challenges, not just for CHF detection. Nevertheless, for experimental data, the Spectrotemporal model produces substantially minor loss values. In Li, et al. [12], an efficient Support Vector Machines (SVM) Fast Conditional Mutual Information (FCMIM)-based detection method for cardiac disease is established. The experimental outcomes specify that the presented algorithm (FCMIM) is able to be used in conjunction with an SVM classifier to develop an efficient system for identifying heart-related problems. In contrast to previously proposed approaches, the suggested finding system (FCMIM-SVM) gained good precision. Additionally, the suggested technique is simple to use in healthcare for the detection of cardiac disease. However, infection control and treatment are important after diagnosis. In this study Cenitta, et al. [13] concentrate on improving the selection of the best features and lowering the quantity of features while still making better selections. A developed Squirrel Search Optimization (SSO) technique is provided here to seek out significant elements of heart disease. The proposed Ischemic Heart Disease state-of-the-art optimization method with the Random Forest classifier (IHSSO) model offers improved feature selection with

superior precision in comparison to more extensive state-of-the-art optimization strategies. However, there is a need for improvement with regard to convergence precision and speed of convergence in the present research. The reference [2] offers a unique approach for identifying key features using ML techniques, which increases the precision of cardiovascular disease diagnosis. The hybrid random forest method with a linear model (HRFLM) forecasting model for heart disease results in an enhanced proficiency level with outstanding precision. However, methodologies for choosing characteristics need to be developed to obtain a deeper understanding of the critical variables and improve the accuracy of heart disease prediction. In reference to Ganie, et al. [14], a novel ensemble Quine McCluskey Binary Classifier (QMBC) method for distinguishing individuals with and without cardiac disease is proposed. This QMBC model outperforms any current modern models and previously proposed ways by various academics. Nevertheless, it is critical to determine whether a dataset is balanced and to use resampling procedures as essential. In Ganie, et al. [14], proposed a CNN method for determining advanced heart disease risk by applying structured data. This model achieves a precision of 88%. However, it needs to be expanded to include unstructured data as well. The authors of Shankar, et al. [15] presented a novel method namely enhanced deep learning with CNN (EDCNN) to aid in improving patient forecasting the disease in heart. The diagnostic approach that has been developed demonstrates the ability to evaluate the degree of probability of heart disease as established by the study. According to the evaluation outcomes, an adaptable architecture and subsequent tuning of EDCNN hyper parameters are capable of up to 99.2% precision. However, advanced artificial intelligence is essential for integration in order to enhance accuracy even further.

Methods based on clustering analysis can offer a nonparametric, unsupervised way for data analysis. A unique FCM algorithm is presented, adopting the concept of the Nearest Neighbor (NN) rule [16, 17]. Missing characteristics in medical profiles are unclear and indicated by NN intervals based on closest neighbour data. This paper's main objective is to evolve a Heart Disease Prediction (HDP) with an FCM clustering algorithm using heart disease databases in order to make sensible medical decisions that are not possible with traditional decision support systems. In order to predict heart attack disorders accurately, this research introduces fuzzy c-means clustering techniques [18-21].

The metaheuristic optimization algorithm (MOA), in order to arrive at the best result, primarily imitates natural and human intelligence. Swarm intelligence, evolutionary, physical, human, and chemical-based algorithms are the four basic divisions of MOA. The following algorithms are frequently utilized for feature selection: Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Sine Cosine Algorithm (SCA), Grey Wolf Optimizer (GWO) and others. Here, FCM-based WOA is employed for selecting best feature. A selected feature subset is evaluated using FCM [22-26] while WOA is employed to determine the best subset feature to maximize accuracy. WOA involves pointing the whale in the direction of the best possible answer from any point in space. This algorithm produces continuous vectors with the same dimension and Ishaq, et al. [1] bounds. The purpose of feature selection is accomplished by a binary conversion threshold on the solution vector. ML is related to computational statistics, which utilizes computers to generate forecasts. Data mining is an examination of data, whereas ML focuses on researching and analyzing data through independent learning. ML methods, principles, and implementation domains are among the benefits that computational optimization provide. ML algorithms and approaches will process massive amounts of data and make judgments and predictions [27]. The presented research in Sridevi and Anitha [28] provides a successful coronary artery disease forecasting model for a system to support clinical decisions that combines Synthetic Minority Oversampling Technique Edited Nearest Neighbour (SMOTE-ENN) for balancing data used for training dispersion and for predicting cardiovascular disease. This model uses density-based spatial clustering with noise to detect and eliminate anomalies. As a result, early intervention might need to be implemented to avoid fatalities resulting from delayed heart disease detection. Nevertheless, if a demographic dataset is acquired, opinions from a neighbourhood heart professional need to be supplied in order to validate the dataset and prognosis model.

To implement classification, Support Vector Machine (SVM) is deployed, and it is not biased by outliers. However, it has some disadvantages, like non-applicability for non-linear problems and not being suitable for a large number of features [29-31]. Similarly, Decision Tree (DT) classification exhibits interpretability and does not demand feature scaling. However, DT generates poor results on very small datasets, and hence overfitting can easily occur. To overcome all these issues, the DCNN is proposed in this work, which is utilized to expand the linear model and features for predicting heart diseases.

Therefore, a CNN classification technique is deployed to detect heart disease using Data Mining approach. The FCM clustering approach is used for the initial stage of segmentation to record the patient's heart data. The WOA feature selection method is established for selecting the segmented heart features. The CNN classifier achieves great accuracy in identifying heart illness.

2. PROPOSED WORK

One of the biggest problems faced by society today is heart disease. Manually calculating the probability of developing heart disease based on risk factors is essential. Health problems can be prevented by detecting diseases as early as possible. In this study, DCNN is proposed as a way to enhance the features for obtaining a higher prediction accuracy. Figure 1 shows the diagram for the suggested model.

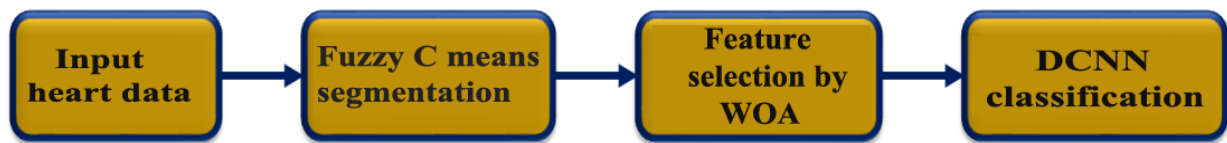


Figure 1. Block diagram for proposed model.

At the first stage, the data is obtained from the input heart data set. Then the data is subjected to segmentation, which is one of the main processes to enhance this approach. The segmented data are then fed to the feature extractor by using the FCM clustering approach, which is used to set up the data into non-overlapping groups. The FCM method extracts the required features by using the WOA algorithm in order to select highly ranked features. A DCNN classifier uses extracted features to classify data and predict the disease without requiring any human intervention. The next part is an overview of the proposed methodology for the detection of heart disease.

2.1. FCM Clustering Technique

In this research, the data set is separated into groups using FCM to maximize intra-group similarity and reduce inter-group variance. Researchers separate the training dataset into various groups before training the Neural Network (NN) prediction model to improve generalization and learning capacity. FCM places data points that belong to various cluster groups with corresponding fuzzy truth values between 0 and 1 using fuzzy partitioning of the dataset. To reduce cluster dissimilarity, the iterative FCM algorithm derives centroids from reciprocal distances in the data. If U is the specification of the dataset matrix, then the FCM initializes the membership matrix as mentioned in Equation 1:

$$\sum_{i=1}^c u_{ij} = 1 \forall j = 1, \dots, n \dots \dots \dots (1)$$

The Equation 2 is defined as dissimilarity function of dataset.

$$J(U, c_1, c_2, c_3, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 (2)$$

Thus, d_{ij} is the Euclidean distance between the i_{th} (c_i) centre and the j_{th} data point, u_{ij} represents the membership of the cluster x_i , with a value between 0 and 1, c_i is the cluster i 's centre and m is the weighting exponent. The centre vector c_i is calculated as follows using an optimization procedure in order to reduce the dissimilarity function:

In addition to do fuzzy partitioning, the objective function described above is iteratively optimized while updating the membership u_{ij} and the cluster centers c_i as follows:

$$\max_{ij} \{u_{ij}^{(k+1)} - u_{ij}^{(k)}\} < \varepsilon \tag{3}$$

When a termination condition between 0 and 1 is met, this iteration will come to an end and the iteration steps are denoted by k . The procedure reaches its end at the saddle point of J_m or a local minimum. The following steps make up the algorithm:

Step 1: Select c, m and ε , when the $\varepsilon > 0$ is a stable positive value: then the initialization of partition matrix is $U^{(0)}, U = [u_{ij}]$

Step 2: while the iteration index is, ($K = 1, 2 \dots$), estimate the matrix prototype $C^{(k)}$ utilizing $C_i, U^{(k)}$. Equation 4 represents the estimated value of C_i .

$$C_i = \frac{\sum_{j=1}^N u_{ij}^m x_i}{\sum_{j=1}^N u_{ij}^m} \tag{4}$$

Step 3: Update values of $U^{(k)}, U^{(k+1)}$

$$u_{ij}^m = 1 / \sum_{k=1}^C (d_{ij} / d_{kj})^{2/(m-1)} \tag{5}$$

Here, Equation 5 presents the updated value of u_{ij}^m .

Step 4: If $\forall i: j: \max_{ij} \{u_{ij}^{(k+1)} - u_{ij}^{(k)}\} < \varepsilon$ then stop and obtain the matrix of cluster prototype C and partition matrix U .

Otherwise set $K = K + 1$ and return step 2

An association between the input parameters and the intended output pattern is made using FCM, which is utilized to reduce the data as part of the segmentation task. FCM is additionally employed to lessen over fitting. Figure 2 illustrates flowchart for FCM clustering technique.

Therefore, segmentation of heart data is established by FCM clustering technique. The following section gives a brief explanation about selection of features from segmented data output by using WOA.

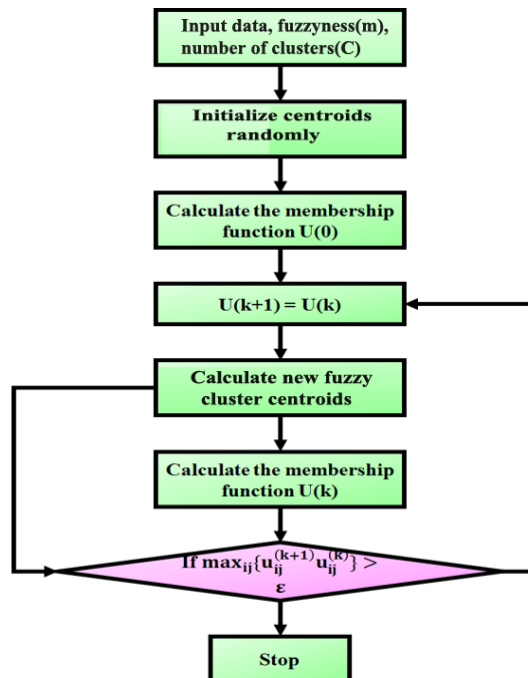


Figure 2. Flowchart for FCM clustering technique.

2.2. Feature Selection by WOA

Humpback whale behaviour in the wild serves as the WOA's main source of inspiration. These whales frequently depend solely on their predatory activity for their survival. Despite the fact that hunting tactics have

previously been employed to address optimisation issues, the whale method is distinctive in that it can employ a random agent in the search space to pursue the prey. Three operators are used in this algorithm's modelling to simulate the exploration, encirclement, and bubble net hunting stages of humpback whales' hunting behaviour. The mathematical formulation is given in the following model and explanation.

2.2.1. Encircling Prey

In this phase, the original best search agent is used by the whale algorithm. The prey's location is extremely near the most effective options available currently. As a result, the remaining agents update the best search agent of their locations. The following sentence serves to express this:

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \tag{6}$$

$$\vec{X}(t + 1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \tag{7}$$

Where current iteration is represented by t and coefficient vectors is denoted by \vec{A} and \vec{C} . Better answer found thus far is represented by the location vector \vec{X}^* and \vec{X} is location vector. The X^* should be modified repeatedly if a better option becomes available. The following formula is used to find out the \vec{A} and \vec{C} vectors:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a}, \tag{8}$$

$$\vec{C} = 2, \vec{r} \tag{9}$$

Where \vec{a} is linear lowered from 2 to 0 throughout number of rounds and r is a random vector in range $[0, 1]$. Through simulations of encircling the prey, this modelling enables any agent to change its location towards the best option. The agents surrounding the most effective solution will aid in the movement in hypercube and it is feasible to explore further into the search space for n dimensions.

2.2.2. Phase of Exploitation

In this stage is also known as bubble net attacking and it employs the following two strategies:

- Shrinkage of the surrounding system: In this phase, of \vec{a} in Equation 3 decreases, resulting in a corresponding drop in the fluctuation range of \vec{A} . This suggests that $[-\vec{a}, \vec{a}]$ was chosen at random to be in. When over the course of optimization, a falls from 2 to 0. According to the unpredictability of \vec{A} in $[-1, 1]$, the search agent's new position can be found between agent's original location and current best location. Figure 3 depicts locations that can be reached by 0 a 1 in a 2-D space from (X^*, Y^*) towards (X, Y) .

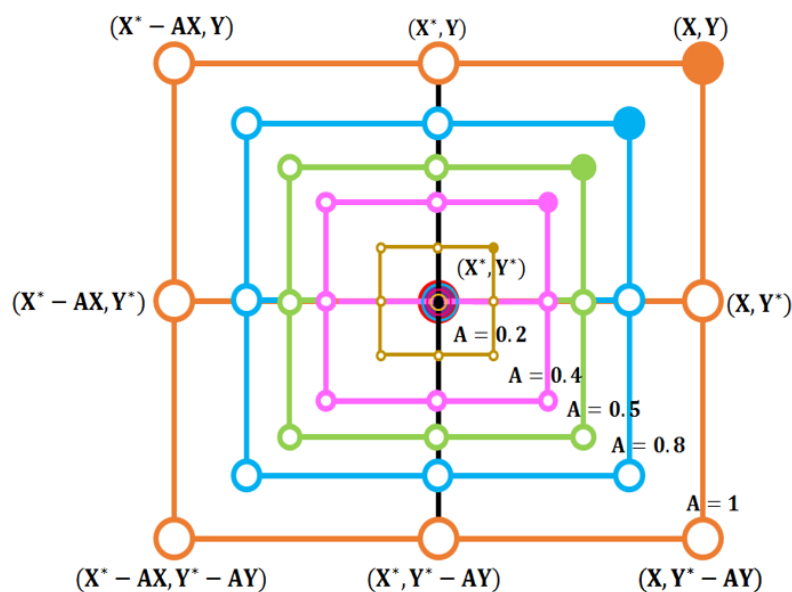


Figure 3. Mechanism of shrinking encircling WOA.

Note: * indicates the solution obtained so far.

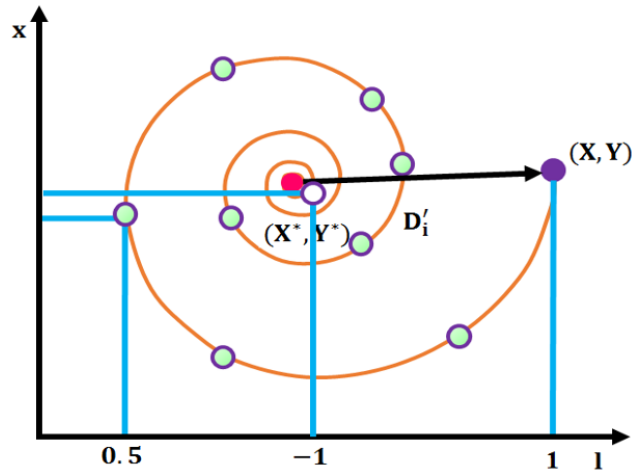


Figure 4. Updating position of WOA spiral.
 Note: * indicates the solution obtained so far.

Figure 4 illustrates the updating position of WOA. This stage involves calculating the distance between whale and its prey. Next, an equation of a spiral is formed between the locations of the whale and its prey to mimic how humpback whales evolved a helix shape. This can be stated as follows:

$$\vec{X}(t + 1) = \vec{D}' . e^{bl} . \cos(2\pi l) + \vec{X}^* (t) \tag{10}$$

Equation 10 displays) where the constant b is utilized to determine logarithmic spiral shape and l is a randomized number between $[-1, 1]$ and distance from i_{th} whale to prey (the best result so far). The whale is simultaneously applying the spiraling, shrinking motion as it moves towards its prey. Next position whale is updated as follows using a 50% assumption that it is possible to transition between the two modes:

$$\begin{cases} \vec{X}(t + 1) = \{\vec{X}^* (t) - \vec{A} . \vec{D}\} \text{ if } p < 0.5 \\ \vec{D}' . e^{bl} . \cos(2\pi l) + \vec{X}^* (t) \text{ if } p \geq 0.5 \end{cases} \tag{11}$$

Where p is a random number in $[0, 1]$.

2.2.3. Phase of Exploration

The WOA achieves worldwide optimization during this phase. In Figure 5, whales randomly arrange themselves in relation to one another as they hunt for prey. The \vec{A} is randomly selected from the range $[-1, 1]$ in order to make search agent deviate from reference whale. This means that A must either be bigger than 1 or less than -1. Furthermore, the WOA performs a worldwide search by selecting a random to update the position of agent.

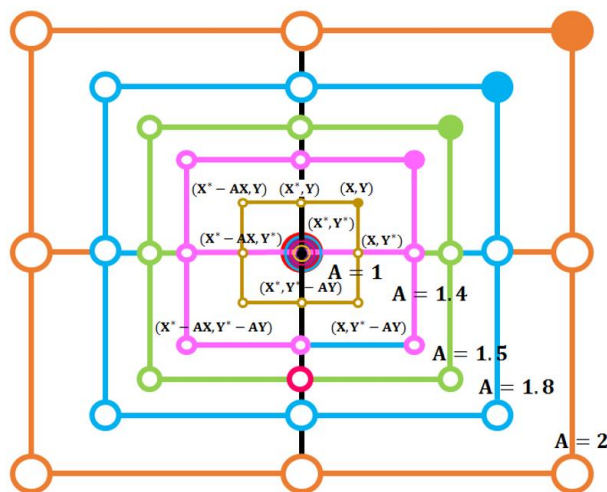


Figure 5. Mechanism of WOA exploration.
 Note: * indicates the solution obtained so far.

The mathematical expression for the modelling of this exploratory method is given as:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (12)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (13)$$

Where, \vec{X}_{rand} is a randomized location for varying whale that were selected from the current population. This paper employs an adaptive WOA to acquire an optimal feature subset that maximizes classification performance. In WOA, the whales constantly change their location in reference to any point in space, starting with the first optimal search agent. Then, as stated in Equations 6 and 7, they attempt to update their locations in reference to the ideal search agent. In the dataset, each particular answer is represented by the same dimension in a continuous vector. The continuous, Ishaq, et al. [1] only values of the solution vector. The values of the solution fitness evaluation have a binary representation. Usually, the number of chosen features and the performance of the classification are matched by the applied fitness function. The equation can be used to express Equation 14.

$$f_{\theta} = \alpha \cdot E + (1 - \alpha) \frac{\sum_i \theta_i}{N} \quad (14)$$

N is a total number of features in dataset and E is the classification error given chosen feature subset and f_{θ} is the fitness function given a vector of size N with 1/0 components describing the selected/unselected feature subset.

The WOA pseudocode is shown in algorithm (1).

```

1) Initialize the whales population  $X_i$  ( $i=1, 2, 3, \dots, n$ ).
2) Compute the fitness of each whale.
3) Set  $X^*$  as the best whale.
4) While ( $t <$  maximum number of iterations) do
for (each search whale) do
Update  $a, A, C, l$  and  $p$ .
If ( $p < 0.5$ ) then
    If ( $|A| < 1$ ) then
The whale position is updating by the
Equation 6.
    else
if ( $|A| \geq 1$ ) then
        Select the random whale  $X_{rand}$ 
        The whale position is updating by
        the Equation 13.
    end
end
    else
if ( $p \geq 0.5$ ) then
        Modify the whale position by the Equation 9.
    end
end
end
    Check if any search goes beyond the search
    space and amend it
    Compute the fitness of each search agent.
    Update  $X^*$  if there is better solution.
     $t=t+1$ 
end

```


In this planned system to implement classification, the DCNN classifier utilized. Which is detailed as follows.

2.3. Classification by DCNN

Figure 6 depicts the architecture of the deep convolutional neural network for mapping protein sequences. It has fourteen layers in total, which include layers of input, flattening, fully connected, convolutional and an output layer. To estimate the likelihood of 1195 data points the softmax function has been implemented in the output layer. The initial layer has input numbers that indicate the location data of a parameter length of the sequence of proteins. In the convolution layer, the filter is employed in the input layer to produce hidden features by batch normalization, convolution, and non-linear change of its components with the activation function. Various window sizes in the one-dimensional convolution layer are evaluated, and two sizes of window have been selected that are near the median length of a protein's beta-sheet and alpha-helix.

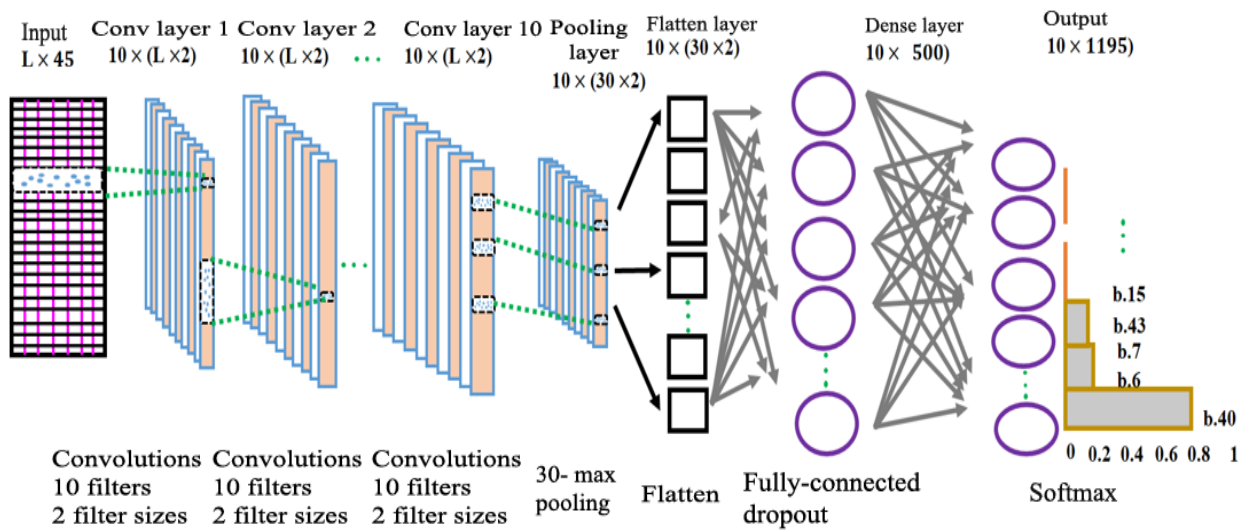


Figure 6. Prediction model architecture of DCNN.

The hidden characteristics produced by the ten filters using two size windows in the initial layer of convolution are used as input for the second layer to be converted in the similar manner. Convolution layer level is set to 10. Based on design, the max pooling layer has been included to turn the variable amount of hidden features in the convolution final layer into a preset amount of features; here, K is fixed to thirty. For each feature map formed by a window size with filter, the 30 highest values are obtained and concatenated. A brief explanation of the various output stages is provided in the following section.

3. RESULTS AND DISCUSSION

Currently, there are a significant number of deaths occurring around the world due to heart attacks. Still, the specialists try to determine the risk factors for heart disease patients. Data mining has been identified as a valuable technique in aiding healthcare professionals in the prediction and anticipation of cardiac disease. As a result, the proposed work uses the DCNN classifier to predict cardiac disease. 1025 samples in total have been trained.

Table 1 demonstrates the input dataset of heart, which comprises a sample of 11 heart data points from people of different ages. The sample data includes sex, age, previous record values, etc...

Table 1. Heart input data.

	Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal	Target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
5	58	0	0	100	248	0	0	122	0	1.0	1	0	2	1
6	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
7	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
8	46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
9	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
10	71	0	0	112	146	0	1	125	0	1.6	1	0	2	1

Table 2. FCM cluster values.

	Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal	Target	Fuzzy
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0	3
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0	3
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0	2
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0	3
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0	0
...
6	59	1	0	140	221	0	1	164	1	0.0	2	0	2	1	3
7	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0	4
8	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0	0
9	50	1	0	110	254	0	0	159	0	0.0	2	0	2	0	4
10	54	0	0	120	188	0	1	113	0	1.4	1	1	3	1	2

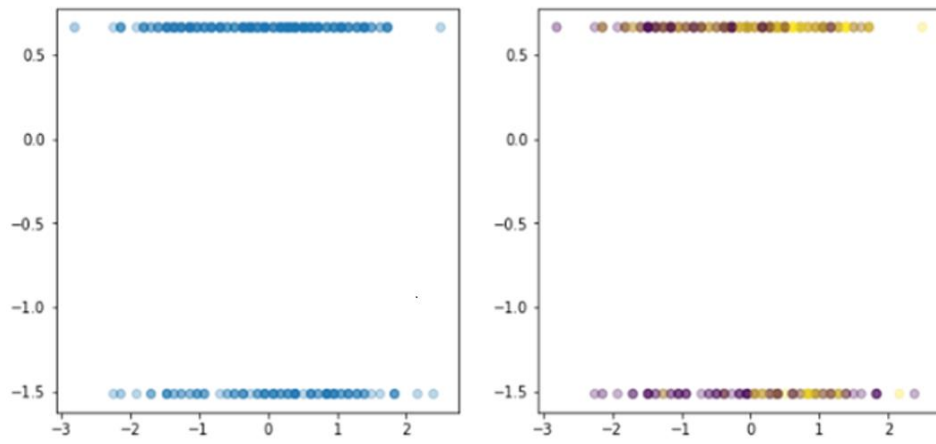


Figure 7. Clustered output.

Table 2 indicates the input data along with the FCM approach. By using the FCM technique, the clustered output attained is illustrated in Figure 7. Clustered features are selected by the WOA algorithm, which is a necessary step in classification.

Table 3. Output parameters.

Output parameters	Values
Accuracy	0.224
Loss	0.901
Specificity	0.840
Sensitivity	0.819

Table 3 shows the outcomes for DCNN classifiers. The proposed work achieves classification accuracy of 90.12% and it is observed that DCNN classifier attains the better specificity and sensitivity with 0.8400 and 81.90%, respectively.

Figure 8 illustrates the diagram for Loss and Accuracy validation of the training data set. Here, the training accuracy increases with number of epochs. Similarly, the loss gets reduced for varying epochs, as indicated in the validation graph.

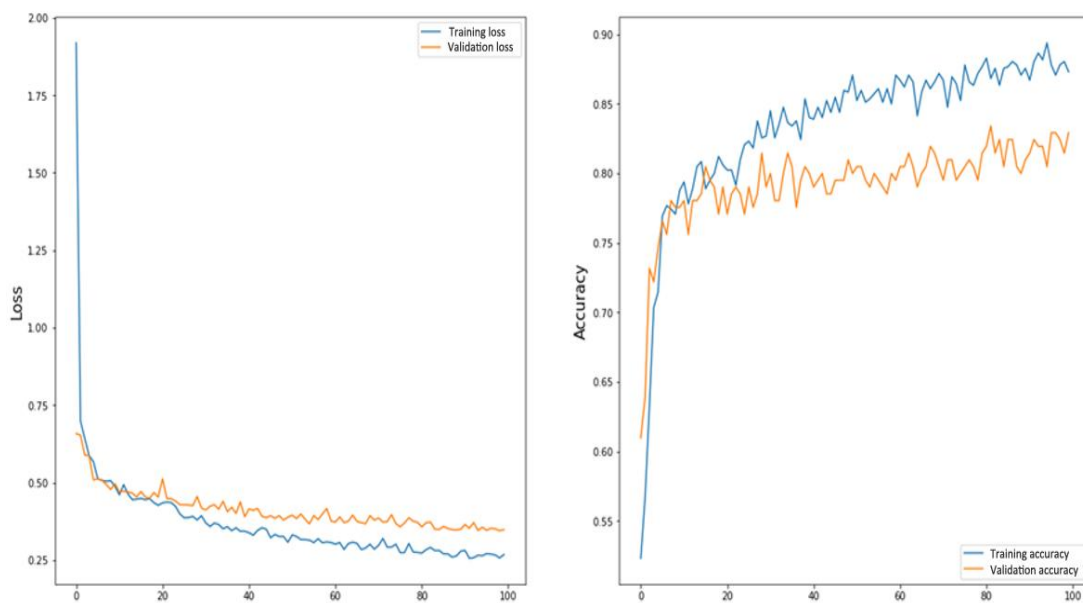


Figure 8. Loss and accuracy validation.

Table 4. Comparison Analysis of K-means and FCM.

Performance metrics	K-means	FCM
Precision	88.09%	91.02%
Recall	87.07%	89.03%

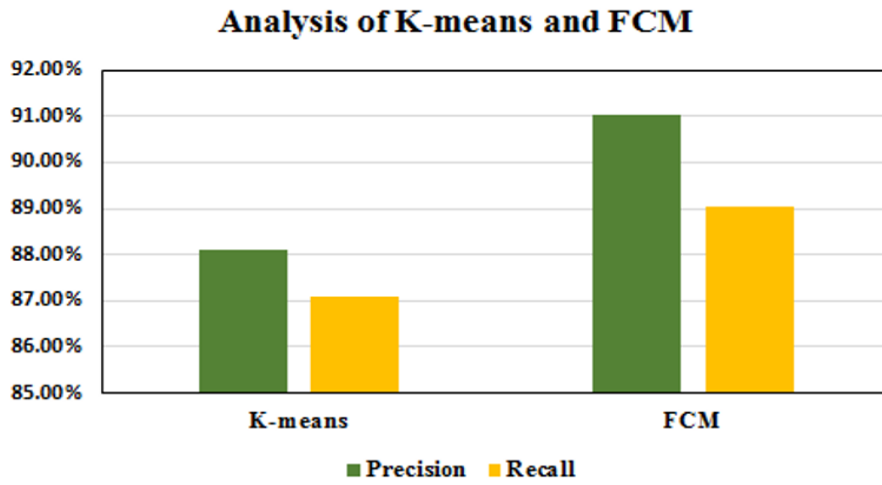


Figure 9. Comparison of K-means and FCM.

The comparison of K-means and FCM is shown in Table 4, and similarly, the relevant plots are shown in Figure 9. The findings lead to the conclusion that FCM works. This method gives significantly better classification accuracy than K-means. Precision and recall are increased through the use of the FCM algorithm.

Figure 10 shows the convergence speed for GA, PSO, GWO, and WOA. From the plot, it is noted that the convergence of WOA is better than that of other algorithms, and the output of the model for WOA can fit the actual output of the system well.

Figure 11 illustrates that the graphical representation of the proposed DCNN classifier with WOA accuracy, specificity and sensitivity is compared with classifiers like SVM, Naïve Bayes, KNN, Logistic Regression, and DCNN with PSO. From the observation, it is clear that the suggested DCNN classifier achieved the enhanced accuracy of 90.12%, specificity of 0.8400 and sensitivity of 81.90%, respectively. Table 5 represents the comparison analysis of the performance metrics values attained for the proposed classifier with existing ones.

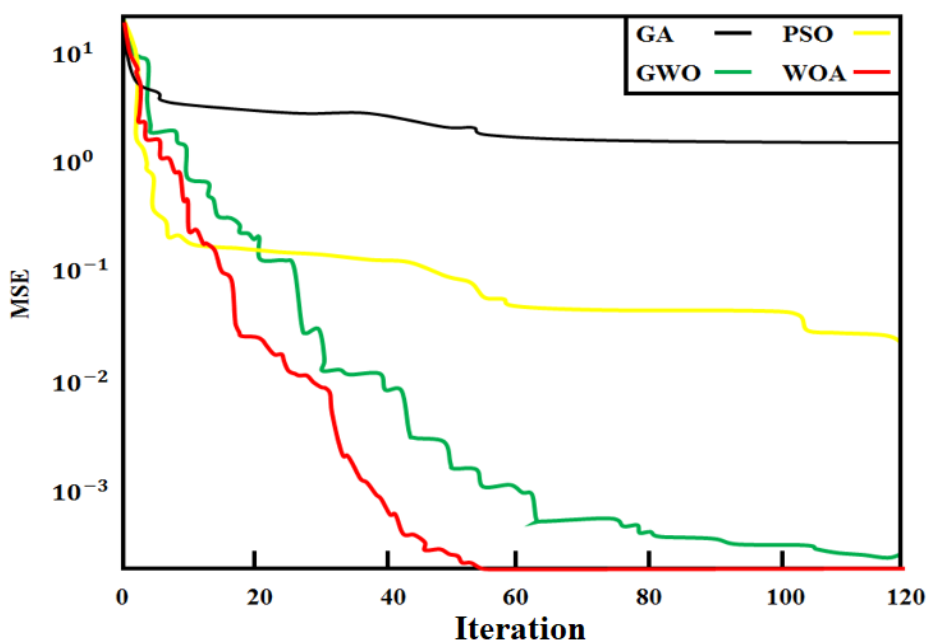


Figure 10. Comparison of convergence speed with WOA.

Table 5. Comparison Analysis performance matrices.

Classifiers	Accuracy	Specificity	Sensitivity
DCNN with WOA	90.12%	84.00%	81.90%
CNN with PSO	85.60%	76.32%	82.42%
Logistic regression	73.66%	70.25%	78.57%
Naïve Bayes	73.17%	71.43%	75.27%
SVM	68.78%	66.39%	72.09%
KNN	65.37%	64.29%	66.67%

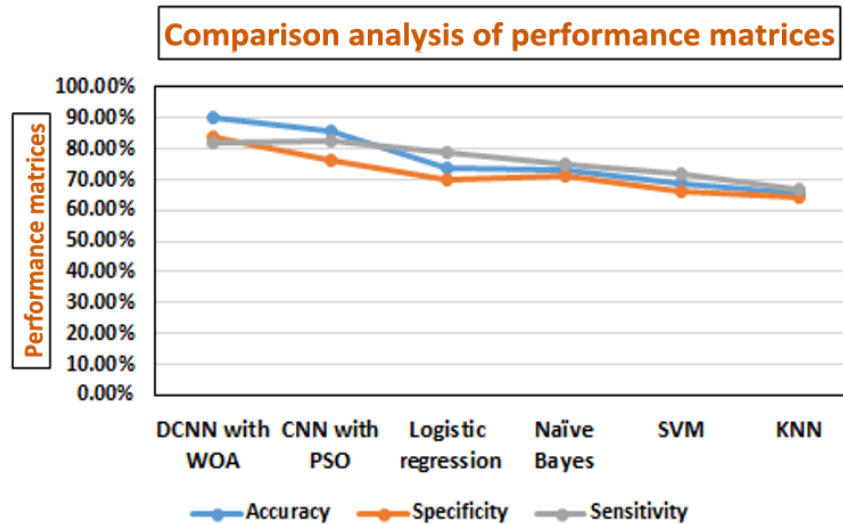


Figure 11. Comparison of accuracy, specificity and sensitivity.

4. CONCLUSION

The prevention of heart attacks may be possible through early detection of heart diseases. A good classification algorithm may enable the doctor to identify the potential for heart disease before it appears. This study employs a DCNN classifier and WOA, along with FCM, to forecast potential heart disease. Following segmentation, the FCM clustering technique is used, assisted by WOA, to select the best feature set. Without any human involvement, the DCNN classifier is deployed to diagnose early-stage heart disease. The attained outcomes demonstrate that the suggested model is implemented in Python software and outperforms other existing approaches in terms of performance, generating an accuracy of 90.12 %, a specificity of 84%, and a sensitivity of 81.90%.

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: Conceptualization and data curation, writing-original draft: R.R.; methodology, writing-review & editing, K.P.K., B.R. and R.R.; project administration, supervision and validation, K.P.K. and B.R. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] A. Ishaq *et al.*, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE Access*, vol. 9, pp. 39707-39716, 2021. <https://doi.org/10.1109/access.2021.3064084>
- [2] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542-81554, 2019. <https://doi.org/10.1109/access.2019.2923707>
- [3] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, p. 104672, 2021. <https://doi.org/10.1016/j.complbiomed.2021.104672>

- [4] J. H. Joloudari *et al.*, "FCM-DNN: Diagnosing coronary artery disease by deep accuracy fuzzy C-means clustering model," *Mathematical Biosciences and Engineering*, vol. 19, no. 4, pp. 3609-3635, 2022. <https://doi.org/10.3934/mbe.2022167>
- [5] A. I. Shima'a Ouf, "A proposed paradigm for intelligent heart disease prediction system using data mining techniques," *Journal of Southwest Jiaotong University*, vol. 56, no. 4, pp. 220-240, 2021. <https://doi.org/10.35741/issn.0258-2724.56.4.19>
- [6] O. Embarak and M. Tarawneh, "Hybrid approach for heart disease prediction using data mining techniques," advance in internet, data and web technologys," presented at the The 7th International Conference on Emerging Internet, Data and Web Technology Springer International Publishing, 2019.
- [7] M. Aljanabi, M. H. Qutqut, and M. Hijawi, "Machine learning classification techniques for heart disease prediction: A review," *International Journal of Engineering & Technology*, vol. 7, no. 4, pp. 5373-5379, 2018.
- [8] S. Rao *et al.*, "An explainable transformer-based deep learning model for the prediction of incident heart failure," *Lee Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 3362-3372, 2022.
- [9] T. Amarbayasgalan, V.-H. Pham, N. Theera-Umpon, Y. Piao, and K. H. Ryu, "An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets," *IEEE Access*, vol. 9, pp. 135210-135223, 2021.
- [10] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36955-36967, 2021.
- [11] M. Gjoreski, A. Gradišek, B. Budna, M. Gams, and G. Poglajen, "Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds," *IEEE Access*, vol. 8, pp. 20313-20324, 2020. <https://doi.org/10.1109/access.2020.2968900>
- [12] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562-107582, 2020. <https://doi.org/10.1109/access.2020.3001149>
- [13] D. Cenitta, R. V. Arjunan, and K. Prema, "Ischemic heart disease prediction using optimized squirrel search feature selection algorithm," *IEEE Access*, vol. 10, pp. 122995-123006, 2022. <https://doi.org/10.1109/access.2022.3223429>
- [14] S. M. Ganie, M. B. Malik, and T. Arif, "Machine learning techniques for big data analytics in healthcare: Current scenario and future prospects in telemedicine: The computer transformation of healthcare." Cham: Springer International Publishing, 2022, pp. 103-123.
- [15] V. Shankar, V. Kumar, U. Devagade, V. Karanth, and K. Rohitaksha, "Heart disease prediction using CNN algorithm," *SN Computer Science*, vol. 1, no. 3, p. 170, 2020.
- [16] Y. Pan, M. Fu, B. Cheng, X. Tao, and J. Guo, "Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform," *IEEE Access*, vol. 8, pp. 189503-189512, 2020.
- [17] R. Kapila, T. Ragunathan, S. Saleti, T. J. Lakshmi, and M. W. Ahmad, "Heart disease prediction using novel quine mccluskey binary classifier," *IEEE Access*, vol. 11, pp. 64324-64347, 2023. <https://doi.org/10.1109/access.2023.3289584>
- [18] V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *Journal of Algorithms & Computational Technology*, vol. 12, no. 2, pp. 119-126, 2018. <https://doi.org/10.2139/ssrn.3139141>
- [19] R. Singh and E. Rajesh, "Prediction of heart disease by clustering and classification techniques prediction of heart disease by clustering and classification techniques," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 5, pp. 861-866, 2019. <https://doi.org/10.26438/ijcse/v7i5.861866>

- [20] N. Yathiraju, A. Sankar, S. Sandhiya, and K. Santhosh, "Cardiac disease prediction for heart monitoring using data mining techniques," presented at the In 2022 International Interdisciplinary Humanitarian Conference for Sustainability IEEE, 2022.
- [21] M. K. Islam, M. S. Ali, M. S. Miah, M. M. Rahman, M. S. Alam, and M. A. Hossain, "Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm," *Machine Learning with Applications*, vol. 5, pp. 1-8, 2021. <https://doi.org/10.1016/j.mlwa.2021.100044>
- [22] H. Alkeshuosh, "Using PSO algorithm for producing best rules in diagnosis of heart disease," presented at the International Conference on Computer and Applications, 2017.
- [23] M. Bahaj and Y. Khourdifi, "Heart disease prediction and classification using machine learning algorithm optimize by particle swarm optimization and ant colony optimization," *The International Journal of Inventive Engineering and Sciences*, vol. 121, pp. 242–252, 2019. <https://doi.org/10.46501/ijmtst061282>
- [24] C. B. Gokulnath and S. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease," *Cluster Computing*, vol. 22, no. S6, pp. 14777-14787, 2019. <https://doi.org/10.1007/s10586-018-2416-4>
- [25] F. Ali *et al.*, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208-222, 2020. <https://doi.org/10.1016/j.inffus.2020.06.008>
- [26] P. T. Sheeba, D. Roy, and M. H. Syed, "A metaheuristic-enabled training system for ensemble classification technique for heart disease prediction," *Advances in Engineering Software*, vol. 174, p. 103297, 2022. <https://doi.org/10.1016/j.advengsoft.2022.103297>
- [27] Y. Shrinkhala, "Heart disease prediction using machine learning techniques," presented at the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, IEEE, 2020.
- [28] N. Sridevi and S. Anitha, "Heart disease prediction using data mining techniques," *Journal of Analysis and Computation*, vol. 13, no. 2, pp. 48-55, 2019.
- [29] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, pp. 133034-133050, 2020. <https://doi.org/10.1109/access.2020.3010511>
- [30] M. Tejamma, P. Naveenkumar Jayakumar, and S. Patil, "A model based on convolutional neural network to predict heart disease," *Journal of Algebraic Statistic*, vol. 136, pp. 2360–2367, 2022.
- [31] A. A. Samir, A. R. Rashwan, K. M. Sallam, R. K. Chakraborty, M. J. Ryan, and A. A. Abohany, "Evolutionary algorithm-based convolutional neural network for predicting heart diseases," *Computers & Industrial Engineering*, vol. 161, p. 107651, 2021. <https://doi.org/10.1016/j.cie.2021.107651>

Views and opinions expressed in this article are the views and opinions of the author(s), Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.