ARTICLE

Heart Disease Prediction Using Convolutional Neural Network with Elephant Herding Optimization

P. Nandakumar and R. Subhashini*

School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, 632014, India
*Corresponding Author: R. Subhashini. Email: rsubhashini@vit.ac.in

ABSTRACT

Heart disease is a major cause of death for many people in the world. Each year the death rate of people affected with heart disease increased a lot. Machine learning models have been widely used for the prediction of heart disease from the different University of California Irvine (UCI) Machine Learning Repositories. But, due to certain data, it predicts less accurately, whereas, for large data, its sub-model deep learning is used. Our literature work has identified that only traditional methods are used for the prediction of heart disease. It will produce less accuracy. To produce more efficacy, Euclidean Distance was used in this work for data pre-processing that will clean the unwanted data and metaheuristics bio-inspired algorithm such as elephant herding optimization (EHO) is utilized for feature selection. Then, this article proposes deep learning models such as convolutional neural network (CNN) and Inception-ResNet-v2 model for the prediction of heart disease from the benchmark dataset such as the UCI Cleveland heart dataset. Finally, the proposed hybrid model utilizes a convolutional neural network with an Inception-ResNet-v2 in the third layer of the architecture that classifies heart disease with the promising result of 98.77%, accuracy for the Cleveland dataset which outperforms all the other state-of-the-art methods. In future work, this model can be used to predict other diseases such as cancer, brain tumor and COVID-19 in available datasets for the betterment of human lives.

KEYWORDS

Heart disease; deep learning; convolutional neural network; elephant herding optimization; Inception-ResNet-v2; prediction

1 Introduction

The heart, which controls blood flow, is a vital component of the human body. All the body's organs receive oxygen and energy from it as well. An abnormal blood flow in the body brought on by a heart infection can be fatal to human life. There are two primary groups of risk factors for heart infections. One is based on a person's family history, age, sex and the other is based on their habits, such as smoking, drinking alcohol and being physically inactive for an extended period. Heart disease has significantly increased the death rate during the last 20 years. Cardiomyopathy, coronary artery disease, heart valve disease, heart arrhythmias, heart failure and congenital heart disease are the various forms of heart disease. According to WHO data [1], heart disease affects 17.9 million individuals worldwide

and is the most common non-communicable disease. According to recent statistics, China is the top nation with a high death rate among those who have heart disease in 2019. Due to their internal organ failures, heart disease [2] is also elevated to the leading complication for other human-killing diseases. Additionally, people of various ages, from 30 to 70, are impacted by it. For many people around the world, early detection and prediction of heart disease are essential. Around 19 million deaths worldwide in 2020 were linked to heart disease, an increase of 18.7% from 2010. Heart attacks are significant medical conditions that occur when the blood supply to the heart muscle is suddenly interrupted and damages the heart muscle. It used to be quite rare for someone under the age of 40 to get a heart attack, but now people under 40 make up one in every five heart attack patients. Another unsettling data to highlight the problem is as follows: A heart attack is more likely to occur in your 20s or early 30s. The number of heart attacks among people in this young age range increased by 2% year between 2000 and 2016. Many research findings show that features or attributes with continuous data are used for the prediction or detection of heart disease. But, to ensure good accuracy in predicting it, many machine learning and deep learning models have been utilized. In this work, the following datasets are considered for the analysis of heart disease from UCI Cleveland.

Machine learning [3] is a mechanism of learning itself from a certain experience. It aims to mimic human intelligence with computer machines. Based on the input it will do the task and produce the output. It can be used in several applications such as image recognition, speech recognition and prediction of human-related issues such as traffic, healthcare, climate and natural disasters. Out of these applications, our major area is healthcare prediction. It is a necessary one nowadays due to a lot of new diseases spreading and infecting many numbers of people around the globe. Machine learning is used in a variety of applications for the prediction of diseases such as heart, cancer and kidney diseases. Diseases can be split into two ways communicable and non-communicable diseases. Machine learning model accuracy is satisfactory for some datasets but not for other large data sets in the medicinal arena. So, the deep learning model is now used widely for the applications utilized by machine learning and AI.

A branch of machine learning called "deep learning" [4] has become well-known worldwide, especially in the field of disease classification. The history of deep learning models was created with the powerful neural network architecture which has a neuron concept and the first invented model is perceptron by Frank Rosenblatt in the year 1957. The neural network has several types, feedforward neural network, recurrent neural network, radial basis function neural network and modular neural network. Deep learning improved a lot with the support of the neural network, so it is now a promising model for many businesses field. It is also a major model for the healthcare sector with more potential to improve human lives with more accurate prediction, detection and diagnosis of various diseases such as heart, cancer and kidney. This paper, mainly focused on major human diseases such as heart disease prediction using machine learning and deep learning.

Overfitting and underfitting are common issues when creating prediction models using existing machine learning methods. Overfitting happens when a model is overtrained and overly complex on a limited dataset, leading to strong performance on training data but poor generalization to new data. The likelihood of bias and discrimination is a serious weakness in existing systems. When compared to other machine learning classifiers, neural networks fared better. Some of the data pre-processing that is generally involved with machine learning is eliminated with deep learning. These algorithms can handle text and visual data that is unstructured and automate feature extraction, reducing the need for human specialists. Graphics processing unit (GPU) computers are optimal for Deep Learning's [5] calculations. Parallel-computing operations, including matrix multiplication, activation functions, and convolutions, make GPUs useful for many fundamental DL techniques. The bandwidth is greatly

enhanced with stacked memory and modern GPU models. As a result, we know that deep learning executed on a GPU yields accurate results and greatly simplifies the computing process. The proposed approach is efficiently utilized in healthcare environments.

Recently in various optimization research, metaheuristics bio-inspired algorithm has been widely applied. Metaheuristics bio-inspired algorithm is also used as feature selection that is utilized in the proposed work. The literature survey on various diseases with the existing model in connection with heart disease is also presented in Section 2.

The main contribution of this work is:

- First, the data is pre-processed by the traditional Euclidean Distance method, followed by the feature selection process with the elephant herding optimization (EHO).
- Then, the popular classifier convolutional neural network with an Inception-ResNet-v2 is utilized for the classification of heart disease from the well-known UCI data repository such as the Cleveland dataset.
- The proposed hybrid model has achieved significant classification performance with more robust results when compared with other meta-heuristic and deep learning models.

The following section of this paper continues with a literature survey on existing machine learning and deep learning models used for heart disease classification in Section 2. The proposed hybrid deep learning model for heart disease prediction is described in Section 3. In Section 4, experimental results are discussed with tabulation and diagrams. Finally, Section 5 presents the future scope of predicting heart disease using deep learning embedded hybrid models to achieve optimal accuracy.

2 Related Works

Several studies have documented the development of deep learning and machine learning models with the potential to diagnose cardiac disease to improve performance while making a heart disease prediction. Researchers frequently assess the effectiveness of prediction models using the publicly accessible heart disease datasets Cleveland, Statlog and Framingham. This section reviews recent research that is pertinent to the current topic.

Gudadhe et al. [6] have established a diagnosis system for heart disease identification using a multilayer perceptron and support vector machine model and achieved an accuracy of 80.41%. They used backpropagation for training MLP and classified heart disease into five classes, 0 represents the absence of disease and 1, 2, 3, 4 represent four types of heart disease.

Kahramanli et al. [7] have developed a heart disease profiling model by applying a hybrid artificial neural network with an addition of fuzzy logic and secured a performance of 86.8%. They have used k-fold cross-validation for the classification purpose and additionally, they have tried this model for the diabetes dataset too and attained a performance of 84.24%.

Das et al. [8] have proposed an ensemble-based diagnosis system using an artificial neural network for heart disease and obtained an accuracy of 89.01%. They have used SAS-based software for the diagnosis of heart disease and achieved the metrics of sensitivity and specificity with 80.95% and 95.91%.

Olaniyi et al. [9] have established a three-phase model with an artificial neural network for heart disease detection and attained an accuracy of 88.89%. They have designed an intelligent system with a support vector machine and feed-forward multilayer perceptron for the prevention of misdiagnosis which is the major error done by many doctors.

Liu et al. [10] have developed the heart disease identification model using relief and rough set techniques and got a maximum performance of 92.59% by the jackknife cross-validation scheme. They have done this in two phases, first with data discretization, feature extraction and feature reduction using a heuristic approach. Second, an ensemble classifier of C4.5 is used for classification purposes.

Geweid et al. [11] have proposed a heart disease labeling model by utilizing an advanced support vector machine-based optimization hybrid method with a nonparametric algorithm for training and obtained an accuracy of 94.97%. They have used a dual SVM model to detect heart failure disease from ECG signals data that results in increased reliability and high accuracy for the identification of heart disease.

Khan et al. [12] have developed an IoMT framework for the detection of heart illness combining modified salp swarm optimization (MSSO) and an adaptive neuro-fuzzy inference system (ANFIS) to increase the precision of prediction. They have used the Levy flight method and enhanced the search capabilities. The author's prediction model achieves a higher accuracy of 99.45 when compared to existing methods.

The short survey on heart disease prediction using artificial intelligence models is shown in Table 1.

Table 1: A recent survey on heart disease prediction

Reference	Year	Models	Dataset	Accuracy	Limitations
[13]	2022	GAPSO-RF	Cleveland	95.60%	High computing cost and temporal complexity.
[14]	2021	GA with LDA and ensemble models	Statlog	93.65%	The performance is demonstrated using only the ROC curve and accuracy.
[15]	2021	DT, RF, KNN, MLP and NB	Framingham	91.39%, 92.80%, 92.68%, 92.64% and 90.56%	The authors used traditional data preprocessing methods.

From the existing literature survey on heart disease prediction, much research has been found on machine learning and deep learning with traditional feature selection and extraction methods. Several studies have used artificial neural network (ANN), deep neural network (DNN), random forest (RF), decision tree (DT), support vector machine (SVM), convolutional neural network (CNN) and k-nearest neighbors (KNN) models for the classification. Most study results in convergence and local optima problem in the prediction of heart disease with machine learning and deep learning models. The significance of data quality comes from the fact that effective decision-making is reliant on it. Even minor changes in the dimensions of the data might have a major impact on the information for making judgments. Therefore, it is advantageous for enterprises to use hybrid technologies that improve the results, as well as verified studies on the selection and evaluation of computational learning approach characteristics. Bio-inspired meta-heuristics algorithms have also been used for selecting the optimal features. But still, it results in slow convergence and local optimum. So, from the above survey, this article added a metaheuristics-based algorithm for the selection of optimal features. The planned research employed to predict heart disease is described in more detail in the following section.

3 Proposed Models

Several studies have been done by numerous authors who predict cardiac disease by implementing hybrid deep learning approaches and additional research in the pre-processing stage research is still insufficient to increase the heart disease prediction accuracy rate. To overcome these difficulties and produce the best prediction outcomes, the hybrid model concept which depends on the differential modeling concept of nonlinear and linear components had been introduced. Additionally, it is believed that employing several or hybrid learning algorithms will produce improved prediction and performance when compared to creative learning methods. The hybrid model is the most popular and well-known model for prediction paradigms for the reason mentioned above. This article has proposed hybrid deep-learning models for the prediction of heart disease. This hybrid system is categorized into three steps: data collection, data pre-processing and model construction. Fig. 1 shows the workflow diagram of the proposed model. The dataset description is explained in Section 3.1, which includes the analysis of a popular benchmark dataset used in this study. Euclidean Distance is utilized for data pre-processing. The data is cleaned using this method. A metaheuristic-based algorithm such as elephant herding optimization (EHO) is used for the feature selection. After the identification of the best features, the selected features are passed to the deep learning models for the classification of the output through a convolutional neural network (CNN) and an Inception-ResNet-v2 model.
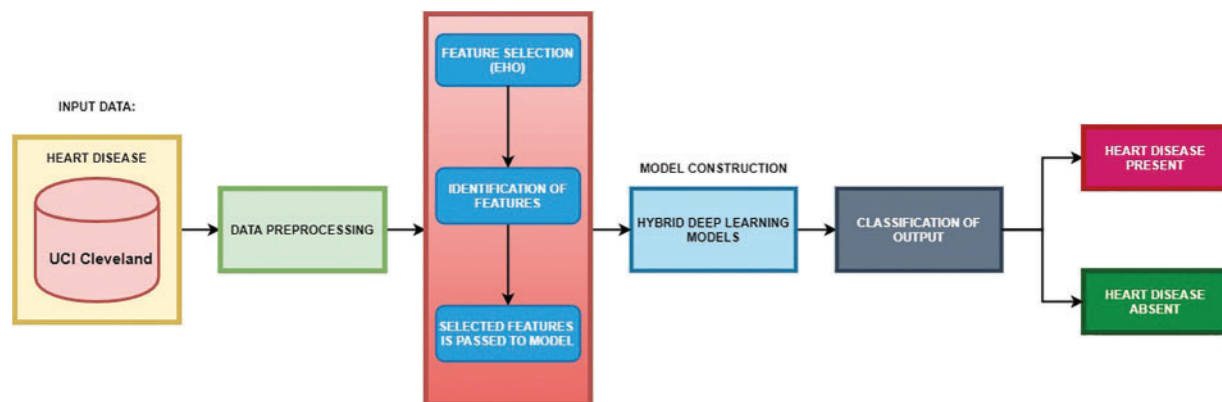

Figure 1: The workflow diagram of the proposed model

3.1 Dataset Description

The proposed model utilized the following three datasets for the prediction of cardiac diseases, such as UCI Cleveland [16]. The Cleveland dataset was taken from the Cleveland clinic foundation. It contains 76 attributes, but the most widely selected features are a subset of 14. Most of the ML researchers used the Cleveland database. The next sections will explain the data pre-processing, feature selection and prediction methods for the presence or absence of heart disease. Table 2 represents the descriptive statistics of the Cleveland dataset values.

3.2 Data Pre-Processing

To improve the quality of the data, pre-processing is used, such as handling missing values and converting the kind of feature using various techniques. The various records shown in the above table of heart disease information are pre-processed and data is cleaned by Euclidean Distance to identify the correct data else it will impact our predictive model. The distance between two points is known as

the Euclidean Distance. The length of a segment connecting the two places is calculated to locate the two spots on a plane. It provides the straight-line distance or the separation between two places. In a two-dimensional plane, let us say that (x_1, y_1) and (x_2, y_2) are two points. The Euclidean Distance equation is then provided by

$$d = \sqrt{\left[(x_2 - x_1)^2 + (y_2 - y_1)^2\right]} \tag{1}$$

where (x_1, y_1) are the coordinates of one point and (x_2, y_2) are the coordinates of the other points. The value stored in the parameter d is the distance between (x_1, y_1) and (x_2, y_2). For instance, the UCI Cleveland dataset initially had 303 patient records; however, 6 of those records contained information that was missing and as a result, unnecessary information was eliminated from the dataset. The enduring 297 patient records have so undergone pre-processing. The multiclass variable and double classification are the features of the supplied dataset. The next section explains the feature selection concept using elephant herding optimization.

Table 2: The visual representation of the Cleveland dataset values

Attributes	Type	Distinct	Minimum	Maximum	Mean	StdDev
Age	Numeric	41	29	77	54.542	9.05
Sex	Numeric	2	0	1	0.677	0.468
Cp	Numeric	4	0	3	2.158	0.965
Trestbps	Numeric	50	94	200	131.694	17.763
Chol	Numeric	152	126	564	247.35	51.998
Fbs	Numeric	2	0	1	0.145	0.352
Restecg	Numeric	3	0	2	0.997	0.995
Thalach	Numeric	91	71	202	149.599	22.942
Exang	Numeric	2	0	1	0.327	0.47
Oldpeak	Numeric	40	0	6.2	1.056	1.166
Slope	Numeric	3	0	2	0.603	0.618
Ca	Numeric	4	0	3	0.677	0.939
Thal	Numeric	3	0	2	0.835	0.957
Condition	Numeric	2	0	1	0.461	0.499

3.3 Feature Selection by Elephant Herding Optimization

In this article, elephant herding optimization (EHO) [17] was used for the feature selection task. In machine learning, feature selection is seen as a pre-processing step. Selecting the most relevant feature subset from a complicated or huge dataset is one of the most difficult challenges. Finding hidden patterns or important knowledge in enormous amounts of data has become a critical issue. It has been proved that feature selection successfully removes unnecessary features. It can also increase the performance of classifiers, lower the computational cost and minimize the amount of storage required. Data has grown in recent years, in terms of both the number of attributes/features and the number of occurrences. Feature selection has shown to be useful in medical categorization and diagnostic aid. An important populace-based technique known as elephant herding optimization (EHO) was developed to solve complex optimization problems. The elephants in each group are informed by their

current location and matriarch through group notifying the operator in this technique. It is created as a result of the disentangling handler's success, which can improve the population variety during the subsequent search phase. With some methodological improvements, EHO keeps exploitation and exploration in synchronization effectively. It is frequently utilized in applications such as energy-based localization, distributed systems, static drone deployment, wireless sensor networks and multilevel image thresholding. The mathematical models of EHO start with the two basic ideas and then introduce the velocity strategy and separation strategy and apply the elitism strategy to the complete algorithm. Each elephant continues the clan update operator according to its current location in the herd and the location of the matriarch.

As a result, for the elephant j in clan ci, the position can be updated as:

$$X_{new,ci,j} = X_{ci,j} + \alpha \times \left(X_{best,ci} - X_{ci,j}\right) \times r, \tag{2}$$

where $X_{new,ci,j}$ and $X_{ci,j}$ are new and old positions for elephant j in clan ci, respectively. $\alpha \in [0, 1]$ is a scale factor. $X_{best,ci}$ represents the best position in clan ci. $r \in [0, 1]$ is a commonly distributed random number.

In Eq. (2), the matriarch's position has not changed. It can be updated as follows for the fittest individual:

$$X_{new,ci,j} = \beta \times X_{center,ci}, \tag{3}$$

$$X_{center,ci} = \frac{1}{n_{ci}} \times \sum_{j=1}^{n_{ci}} X_{ci,j}, \tag{4}$$

where $\beta \in [0, 1]$ is a scale factor. $X_{center,ci}$ is the middle place in clan ci. n_{ci} is the number of elephants in clan ci. The positions of all the clan members may be observed to be modified by the matriarch.

The male elephants live alone after being separated from the herd. The worst elephant in each clan is supposed to be replaced by a separation operator in the EHO algorithm. The procedure can be shown in Eq. (5).

$$X_{worst,ci} = X_{min} + (X_{max} - X_{min} + 1) \times r, \tag{5}$$

where $X_{worst,ci}$ symbolizes the lousiest elephant in clan ci. The elephant position's upper and lower boundaries are X_{max} and X_{min}, respectively.

This research explores the hybrid technique, combining EHO with deep learning models. The key contribution of this research is the proposal of an EHO performance. In this hybrid technique, the metaheuristic bio-inspired methodology is used to find the best feature subset that increases heart disease classification accuracy while reducing feature subset length. The flowchart representation of EHO is shown in Fig. 2.

3.4 Convolutional Neural Network

A convolutional neural network (CNN), often termed as ConvNet, appears to be the best framework for medical image processing. The architecture of a CNN can help with multilayer deep hierarchical learning. CNN's early layers may retrieve minimal data whereas the deep layers could retrieve increased feature representations, which are then merged to reliably pinpoint significant locations. The neurons in the ConvNet layers are structured in three measurements: height, breadth and depth. The fundamental CNN architecture is shown in Fig. 3. The CNN architecture is made up of an input layer, a convolutional layer, a rectified linear unit (ReLu), a pooling layer, a fully connected layer and an output layer.
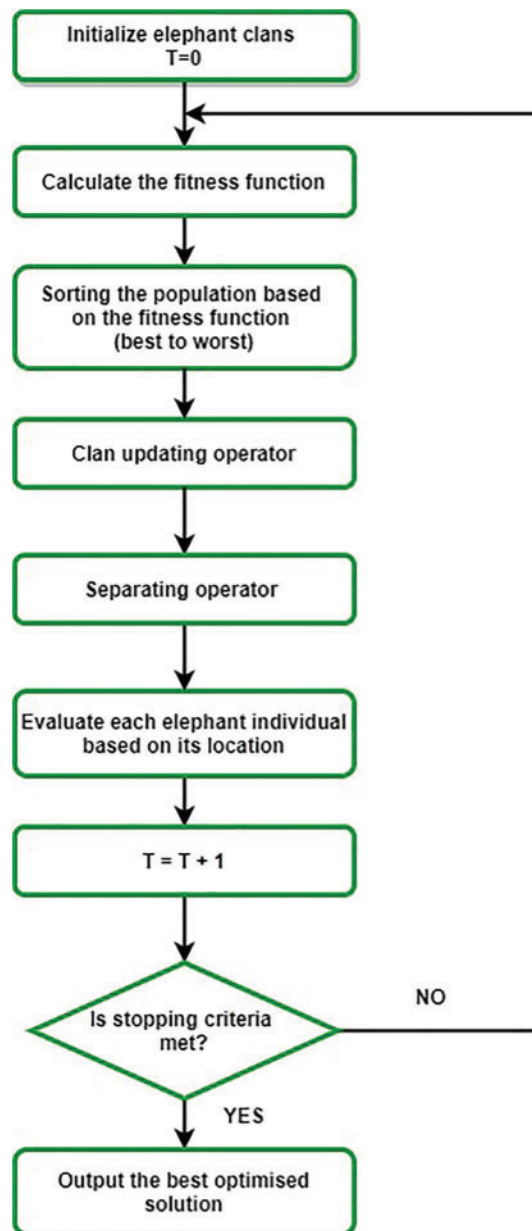
Figure 2: Flow chart of elephant herding optimization algorithm

To create a comprehensive CNN architecture, several layers are piled one on top of the other. The convolutional, ReLU and pooling layers are used to extract features. In the fully linked layer, classification takes place. In a deep neural network, identifying individual layers and associated connections is more difficult than executing a pre-trained general structure. The identification of the connection between layers can be enhanced by building a final neural network model. The model grows better perceptive to emerging themes and works properly in a much more substantial way since it takes out extracted attributes. The early detection of cardiac illness may be made easier with the help of the neural network. The use of pre-trained neural network models for handling, processing,

fragmentation and categorization of radiological pictures of heart disease-affected individuals has been made easier because of transfer learning. Additionally, several neural networks have grown in response to the current task.
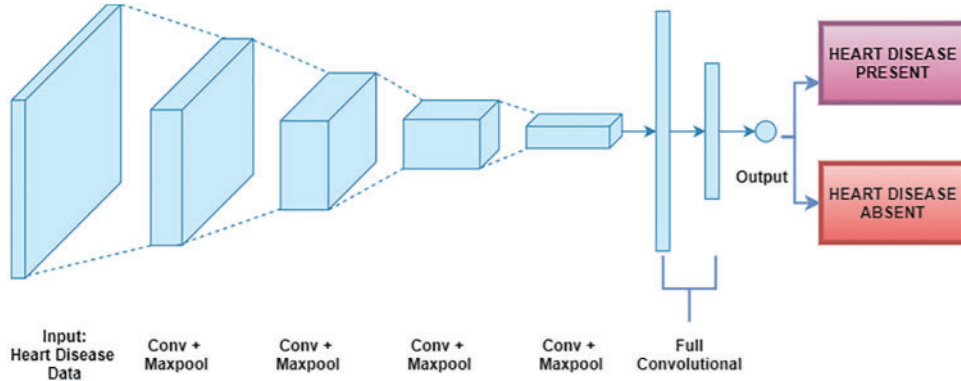

Figure 3: CNN architecture

In this work, the input consists of one-dimensional data with $x = (x_1, x_2, x_3, \ldots, x_{n-1}, x_n, clabel)$ where $x_n \in R^d$ denotes the heart disease features and *clabel* $\in R$ denotes a class label used for the output of either heart disease present or absent. A Conv1D is used for the construction of a feature map Fm. Then, it is applied for the convolution operation on the heart disease input data with filtering of $w \in R^{Fd}$ where F denotes the inherent features of the input data that will produce the final output after feeding it in the next input block.

From the set of features, the new feature map Fm is obtained as follows:

$$h\ell_i^{Fm} = tanh\left(w^{Fm}x_{i:i+F-1} + b\right) \tag{6}$$

where hl is the filter employed for each set of heart disease input features F is defined as

$$\{x_{1:F}, \ x_{2:F+1}, \ x_{3:F+2}, \ \ldots, x_{n-F+1}\} \tag{7}$$

From (7), the generated feature map is

$$hl = [hl_1, hl_2, hl_3, \ldots, hl_{n-F+1}] \tag{8}$$

where $b \in R$ denotes a bias term and the filter $hl \in R^{n-F+1}$.

3.5 Inception-ResNet-v2 Model

The workflow representation of the Inception-ResNet-v2 model is shown in Fig. 4. A convolutional neural network called Inception-ResNet-v2 is created by combining the Inception structure and the Residual connection. Multiple convolutional filters of various sizes are mixed with residual connections in the Inception-ResNet block. In addition to avoiding the degradation issue brought on by deep structures, the inclusion of residual connections shortens training time. This article has classified the UCI heart disease dataset into normal (absence of heart disease) and abnormal (presence of heart disease) using an Inception-ResNet-v2 model. The data is randomly utilized with 80% for training and 20% for testing activity. In Fig. 4, the evaluation goes on until the assessed output is acquired. The method of training is continued until all the training samples are applied consecutively resulting in a possible minimum value of the loss function. After acceptable training process completion, the model is tested with the leftover 20% of data.
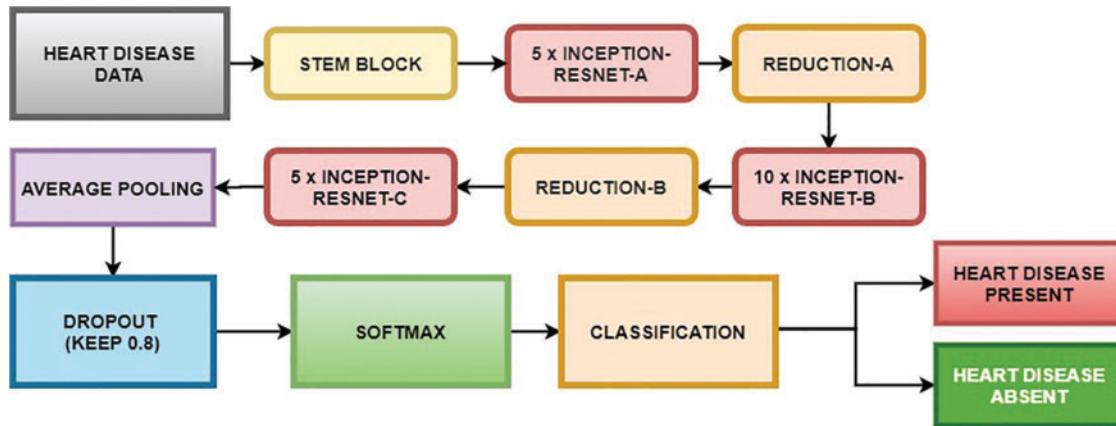
Figure 4: The basic architecture of Inception-ResNet-v2

4 Result and Discussions

This section exhibits the experimental result for the prediction of heart disease using hybrid deep learning models with an optimization method. The proposed hybrid model is shown in Fig. 5. After data pre-processing, from the UCI Cleveland heart disease dataset where the categorical attributes are combined for the recognition of the best features for the accurate prediction of heart disease. The prediction model starts with the input data that is loaded into the machine. The data is pre-processed by Euclidean Distance. Then, it is split into 80–20 percent for training and testing. Then, split the trained data into validation and optimize the parameter using metaheuristics bio-inspired algorithm, Elephant Herding Optimization for feature selection. The next step is to pass the selected features into the trained deep learning models such as CNN with an Inception-ResNet-v2 [18] which used the concept of convolutional layer, max-pooling layer, dropout, IRv2 to the fully connected layer and dense with softmax activation function and classified the output into heart disease present or absent. Then, finally, the prediction of heart disease is measured with performance metrics such as accuracy and time complexity. The correct classification of several cases in the dataset depends on accuracy. The equation used to determine accuracy was represented as

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{9}$$
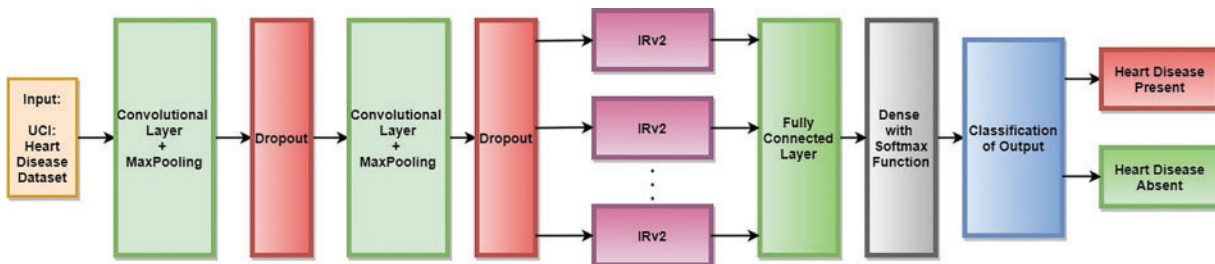

Figure 5: The hybrid deep learning model architecture

This article has suggested an elephant herding optimization approach using a hybrid CNN with an Inception-ResNet-v2 classification model scheme for the categorization of medical abnormalities in heart illness. The EHO technique was used in the suggested scheme to determine the two critical parameters as well as the best feature subset dynamically and concurrently for the hybrid deep learning model. This is the first time, to our knowledge, that the enhanced EHO technique has been used to train the hybrid model. This article has conducted studies on two typical illness diagnosis difficulties to validate the suggested approach and it was then effectively applied to forecast heart disease patients. The experimental findings reveal that the proposed EHO improved the hybrid model's approach that outperforms previous hybrid methods based on meta-heuristics. This means that combining EHO with convolutional neural networks and the Inception-ResNet-v2 model can produce reliable and consistent results for the heart disease categorization issues studied.

4.1 Experimental Results

This section explains the results obtained using the proposed model and a comparison of the existing work for the prediction of heart disease. The categorical data is used in the prediction of heart disease since it is more structured and inherently embed with medical observation. In addition to preprocessing task, this work has also performed K-fold cross-validation to improve the performance of the prediction accuracy. All these records have been utilized with K-fold cross-validation for the implementation part. Table 3 shows the K-fold cross-validation performance of the Cleveland dataset with a K value of 10. The data is divided into 10-folds resulting in 10 subsets, which will use 9-fold for training and 1-fold for testing. Tables 4–6 will show the performance metrics values of the proposed model such as epochs and accuracy with time complexity. Table 4 shows the CNN model performance where it starts with epoch size 109 with an accuracy of 76.45 and ends with epoch size 152 with an accuracy of 97.14, respectively. Table 5 shows the ResNet model performance where it starts with epoch size 219 with an accuracy of 77.85 and ends with epoch size 387 with an accuracy of 97.47, respectively. Table 6 shows the Inception-ResNet-v2 model performance where it starts with epoch size 219 with an accuracy of 91.72 and ends with epoch size 292 with an accuracy of 98.77, respectively. When compared with the time complexity of the proposed work, the Resnet model attains good accuracy at a shorter time at 220 m/s whereas CNN attains a high-performance accuracy at 859 m/s and Inception-ResNet-v2 at 579 m/s. The Big (O) notation is also used to compute the computational complexity. This kind of difficulty is known as time complexity, and this work measures it by accounting for how long it takes our algorithms to run concerning the number of inputs. The proposed model has the time complexity of $O(n)$ where n is the number of inputs. When using smaller datasets to detect heart illness, the time complexity of our suggested model somewhat increased. The proposed model structure is changed since there are fewer instances in the datasets. The intricacy of the algorithm causes a large increase in training time because the new component of the structure needs to be demonstrated again. The complexity and quantity of necessary iterations rise together with the training time. However, the algorithm's temporal complexity somewhat rises as a result of the reduced size of the solution space. The visual representation of the performance metrics for the proposed CNN model is shown in Figs. 6 and 7, respectively. The visual representation of the performance metrics for the proposed ResNet model are shown in Figs. 8 and 9, respectively and finally, the visual representation of the performance metrics for the proposed Inception-ResNet-v2 model are shown in Figs. 10 and 11, respectively.

Table 3: 10-fold cross-validation of the Cleveland dataset

10-fold	Cleveland dataset (297 records)									
	1	2	3	4	5	6	7	8	9	10
1	**1–30**	31–60	61–90	91–120	121–150	151–180	181–210	211–240	241–270	271–297
2	1–30	**31–60**	61–90	91–120	121–150	151–180	181–210	211–240	241–270	271–297
3	1–30	31–60	**61–90**	91–120	121–150	151–180	181–210	211–240	241–270	271–297
4	1–30	31–60	61–90	**91–120**	121–150	151–180	181–210	211–240	241–270	271–297
5	1–30	31–60	61–90	91–120	**121–150**	151–180	181–210	211–240	241–270	271–297
6	1–30	31–60	61–90	91–120	121–150	**151–180**	181–210	211–240	241–270	271–297
7	1–30	31–60	61–90	91–120	121–150	151–180	**181–210**	211–240	241–270	271–297
8	1–30	31–60	61–90	91–120	121–150	151–180	181–210	**211–240**	241–270	271–297
9	1–30	31–60	61–90	91–120	121–150	151–180	181–210	211–240	**241–270**	271–297
10	1–30	31–60	61–90	91–120	121–150	151–180	181–210	211–240	241–270	**271–297**

Table 4: Performance metrics of the proposed CNN model for the Cleveland dataset

Epochs	Accuracy	Time complexity m/s
109/9	76.45	846
112/9	87.43	833
143/2	93.08	832
143/7	95.81	849
144/9	96.34	868
152	97.14	859

Table 5: Performance metrics of the proposed ResNet model for the Cleveland dataset

Epochs	Accuracy	Time complexity m/s
219	77.85	227
227	90.94	224
272	93.11	224
323	94.52	225
342	95.45	221
362	95.91	224
372	97.04	222
387	97.47	220

Table 6: Performance metrics of the proposed Inception-ResNet-v2 model for the Cleveland dataset

Epochs	Accuracy	Time complexity m/s
219/7	91.72	753
224/3	96.20	596
225/6	97.62	584
240/2	98.42	619
245	98.44	628
272	98.65	575
292	98.77	579
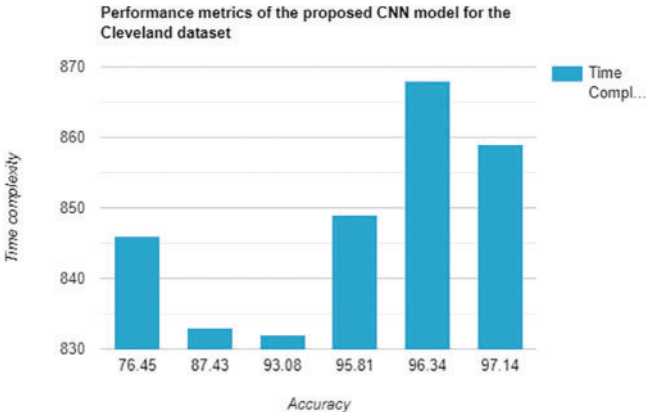
Figure 6: Performance metrics of the CNN model in terms of accuracy *vs.* time complexity
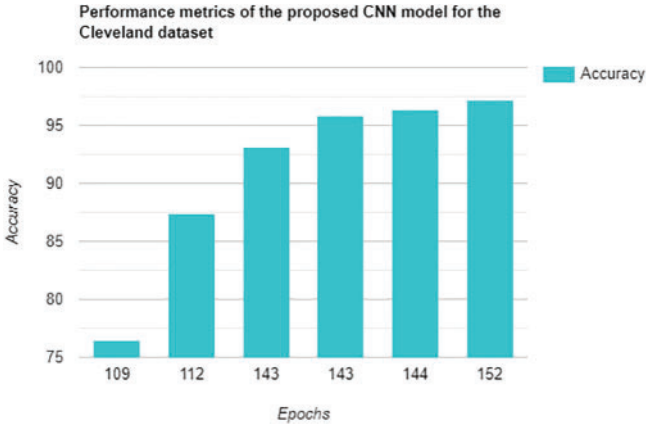
Figure 7: Performance metrics of the CNN model in terms of epoch *vs.* accuracy
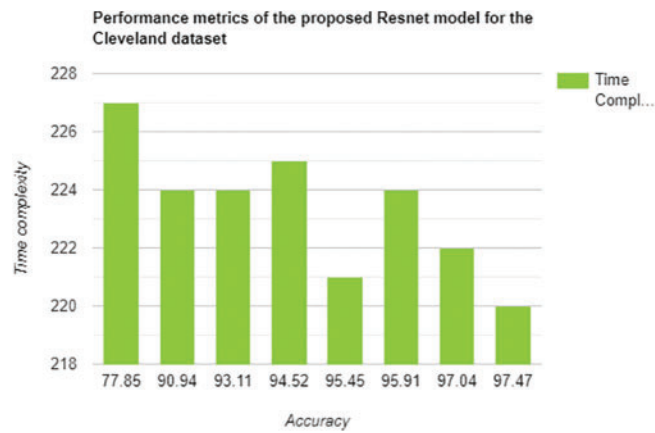
Figure 8: Performance metrics of the ResNet model in terms of accuracy *vs.* time complexity
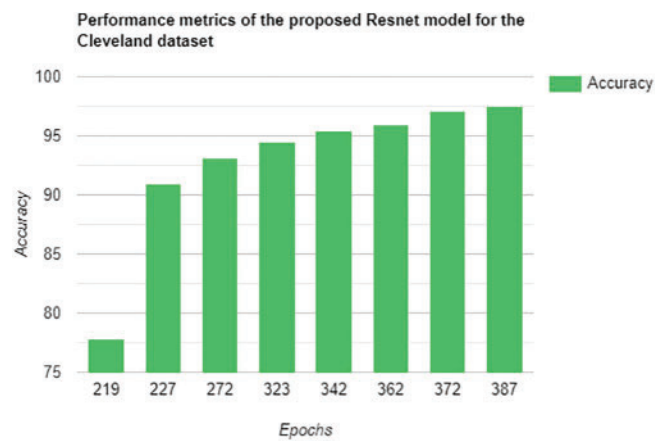

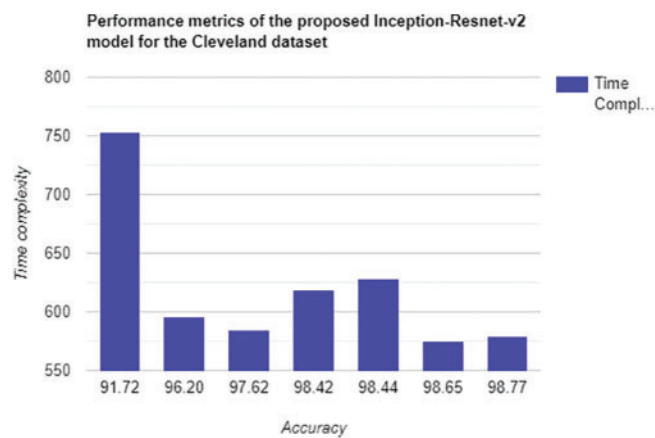**Figure 9:** Performance metrics of the ResNet model in terms of epoch *vs.* accuracy

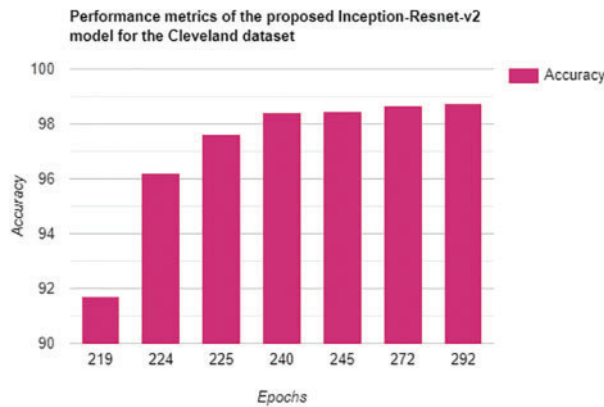**Figure 10:** Performance metrics of Inception-ResNet-v2 model in terms of accuracy *vs.* time complexity

Figure 11: Performance metrics of Inception-ResNet-v2 model in terms of epoch *vs.* accuracy

It is shown that from Table 7 and Fig. 12, the proposed model has achieved better in all the performance metrics such as precision, sensitivity, specificity, F1-score, ROC and accuracy when compared with other models.

Table 7: Performance metrics of the proposed model with other models for the Cleveland dataset

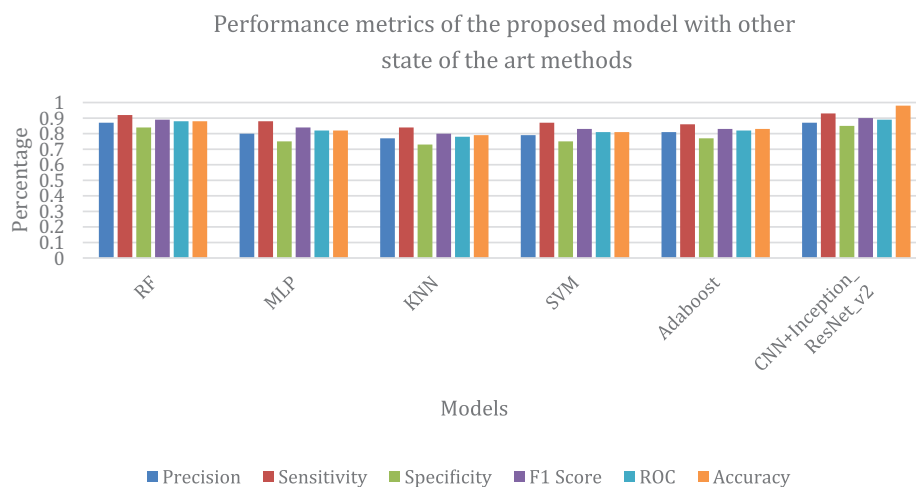| Model	Precision	Sensitivity	Specificity	F1-score	ROC	Accuracy
RF	0.87	0.92	0.84	0.89	0.88	0.88
MLP	0.80	0.88	0.75	0.84	0.82	0.82
KNN	0.77	0.84	0.73	0.80	0.78	0.79
SVM	0.79	0.87	0.75	0.83	0.81	0.81
Adaboost	0.81	0.86	0.77	0.83	0.82	0.83
Proposed	0.87	0.93	0.85	0.90	0.89	0.98

Figure 12: Comparison of the performance metrics of the proposed model in terms of precision, sensitivity, specificity, F1-score, ROC and accuracy

It is shown that from Table 8 and Fig. 13, the proposed model elephant herding optimization and convolutional neural network with the Inception-ResNet-v2 model has achieved better accuracy when compared with other work. From Table 8, Sibo Prasad Patro et al. have designed a framework based on different classifiers with a new feature selection approach of a salp swarm optimized neural network and attained an accuracy of 86.7% in the prediction of heart disease. They have additionally compared their results with other classifiers such as KNN, NB and SVM, respectively. Using the publicly accessible Cleveland Heart Disease dataset, the previous result is compared with several intelligent systems constructed with ML algorithms for predicting whether a person is likely to acquire heart disease. They also discussed the use of a particle swarm optimization (PSO) algorithm to train a multilayer perceptron (MLP) for the diagnosis of heart disease. The algorithms are tested and judged based on their categorization metrics. They have obtained an accuracy of 84.61% with their proposed model MLP-PSO. Most of the existing work is performed with traditional feature selection methods such as principal component analysis, Pearson correlation coefficient, least absolute shrinkage, selection operator, and chi-Squared. In this work, Euclidean Distance and metaheuristics bio-inspired algorithms such as elephant herding optimization were used for the feature selection process. Additionally, this article has utilized the EHO algorithm to perform faster convergence and reduce local optima problems. Earlier work has concentrated on machine learning and deep learning models such as k-nearest neighbors, logistic regression, autoencoder, convolutional neural network, deep neural network, artificial neural network, support vector machine, naïve Bayes and ensemble models. In this work, the hybrid deep learning models were utilized for the prediction of heart disease and it attains better optimal accuracy when compared with other results.

Table 8: Literature of Cleveland dataset with our proposed model (accuracy)

Reference	Year	Model	Accuracy (%)
Murugesan et al. [19]	2021	BPNN	84
Lin et al. [20]	2020	CNN	78.65
Junaid et al. [21]	2020	Hybrid model by combining ANN, SVM, and NB	88
Ali et al. [22]	2021	MI-DNN	93.3
Sandhiya et al. [23]	2020	T-CNN	85.4
Patro et al. [24]	2021	SSA-NN	86.7
Khan [25]	2020	MDCNN	93.3
Al Bataineh et al. [26]	2022	MLP-PSO	84.61
Proposed model	2023	Hybrid model	98.77

4.2 Discussions and Future Work

In this work, the elephant herding optimization is used for the feature selection and optimization for the prediction accuracy of heart disease. Hybrid deep learning models such as convolutional with an Inception-ResNet-v2 are used to predict heart disease presence or absence. The proposed model produces more efficacy when compared with other models. There are some advantages and disadvantages to the proposed work. The advantages are fast convergence and solving local optima problems. The disadvantages are, that the proposed work is executed in existing UCI datasets. The

fact that CNNs need a lot of labelled data to train properly, which can be expensive and time-consuming to gather and annotate, is one of their key drawbacks. Another drawbacks of our proposed Inception-ResNet-v2 model is that, because of the complex layout of the internal modules, training and testing run times are roughly longer than for other models. In the future, this approach can be implemented in the detection of arrhythmia heartbeats [27] where the early work utilized a wavelet transform-based CNN model. Heart rate and blood pressure are also measured from the medical datasets based on the proposed work, where early work [28] utilized a photoplethysmograph signal with the Fourier transform method. Future evaluation and analysis of the suggested hybrid model on a range of medical-based datasets will show its efficacy in medical diagnosis and healthcare. Future studies can also be recommended for optimization issues like parameter tuning and time complexity with other deep learning methods such as artificial neural networks (ANNs), generative adversarial networks (GANs) and radial basis function networks (RBFNs) to experiment on larger datasets with the inclusion of real-time data such as from medical centers and hospitals.
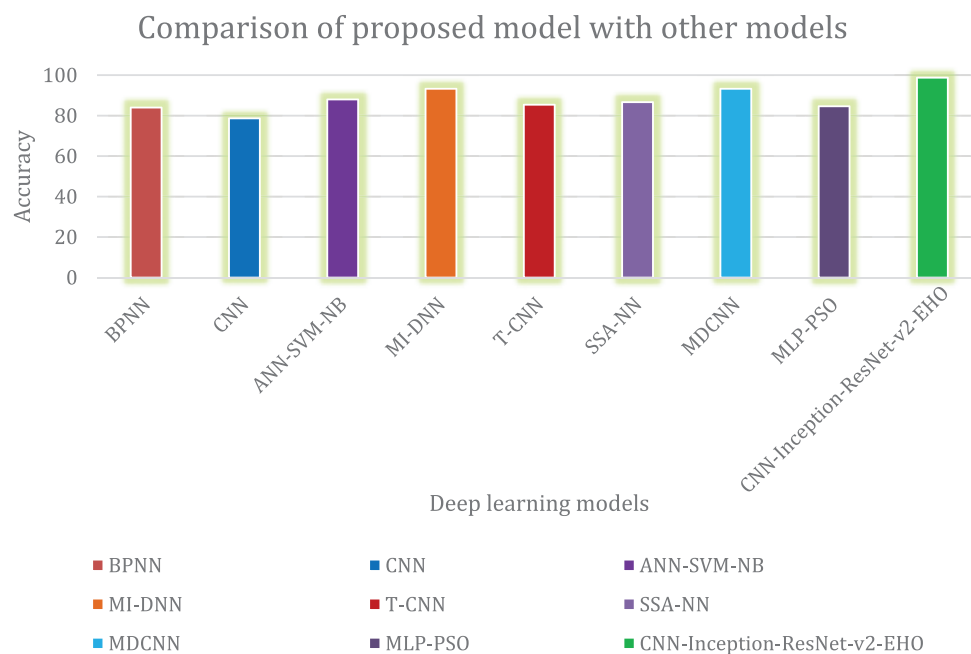

Figure 13: Comparison of the proposed model with other state-of-the-art methods

5 Conclusions

Early heart disease prediction is an important issue for all ages of people. This study aims to improve heart disease prediction accuracy using UCI datasets. This work presents three important methods for the prediction of heart disease, they are elephant herding optimization, convolutional neural network, and Inception-ResNet-v2. The first step is the data pre-processing which is done by the traditional Euclidean Distance method. Next, the input features are transferred to the elephant herding optimization for the selection of the right features that are passed to the classification models. The findings also demonstrate that the elephant herding optimization algorithm improved in terms of classification performance, the number of chosen features and convergence speed. In this work, EHO is also used to minimize the local optimum problem. The proposed convolutional neural network with an Inception-ResNet-v2 model attains an accuracy of 98.77% for the Cleveland dataset which

outperforms other states of the art method. In the future, the proposed model can be added with other new metaheuristics-based methods on large datasets that can be used for the early prediction of heart disease which is a major concern for people after COVID-19 disease. The effectiveness of EHO will be tested on increasingly challenging engineering and scientific topics in the future. Ensemble models can also be used with the proposed model to enhance other performance metrics in the prediction and detection of various human disease problems.

Author Contributions: The authors of this work contributed in the following ways: conceptualization and formal analysis, P.N. and R.S.; methodology, P.N. and R.S.; software, P.N. and R.S.; data curation, P.N.; resources, P.N. and R.S.; supervision, R.S.; writing—original draft, P.N.; writing—review and editing, P.N. and R.S. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in UC Irvine Machine Learning Repository at http://archive.ics.uci.edu/ml/datasets/Heart+Disease.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   W. Tsao Connie, W. Aaron Aday, I. Zaid Almarzooq, A. Alonso, Z. Andrea Beaton *et al.,* "Heart disease and stroke statistics—2022 update: A report from the American Heart Association," *Circulation*, vol. 145, no. 8, pp. 153–639, 2022.

[2]   M. Zuin, G. Rigatelli, C. Bilato, G. Zuliani and L. Roncon, "Heart failure as a complication of COVID-19 infection: Systematic review and meta-analysis," *Acta Cardiologica*, vol. 77, no. 2, pp. 107–113, 2022.

[3]   R. G. Nadakinamani, A. Reyana, S. Kautish, A. S. Vibith, Y. Gupta *et al.,* "Clinical data analysis for prediction of cardiovascular disease using machine learning techniques," *Computational Intelligence and Neuroscience*, vol. 1, no. 1, pp. 1–13, 2022.

[4]   D. Kaul, H. Raju and B. K. Tripathy, "Deep learning in healthcare," in *Deep Learning in Data Analytics.* Cham: Springer, pp. 97–115, 2022.

[5]   M. Nishiga, D. W. Wang, Y. Han, D. B. Lewis and J. C. Wu, "COVID-19 and cardiovascular disease: From basic mechanisms to clinical perspectives," *Nature Review Cardiology*, vol. 17, no. 9, pp. 543–558, 2020.

[6]   M. Gudadhe, K. Wankhade and S. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network," in *Proc. of the Int. Conf. on Computer and Communication Technology (ICCCT)*, Uttar Pradesh, UP, India, IEEE, pp. 741–745, 2010.

[7]   H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Systems with Applications*, vol. 35, no. 1–2, pp. 82–89, 2008.

[8]   R. Das, I. Turkoglu and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.

[9]   E. O. Olaniyi, O. K. Oyedotun and K. Adnan, "Heart diseases diagnosis using neural networks arbitration," *International Journal of Intelligent Systems and Applications*, vol. 7, no. 12, pp. 75–82, 2015.

[10]   X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu *et al.,* "A hybrid classification system for heart disease diagnosis based on the RFRS method," *Computational and Mathematical Methods in Medicine*, vol. 1, no. 1, pp. 1–11, 2017.

[11] G. G. Geweid and M. A. Abdallah, "A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique," *IEEE Access*, vol. 7, pp. 149595–149611, 2019.

[12] M. A. Khan and F. Algarni, "A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS," *IEEE Access*, vol. 8, pp. 122259–122269, 2020.

[13] M. G. El-Shafiey, E. S. A. El-Dahshan, A. Hagag and M. A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 18155–18179, 2022.

[14] V. Jothi Prakash and N. K. Karthikeyan, "Enhanced evolutionary feature selection and ensemble method for cardiovascular disease prediction," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 13, no. 3, pp. 389–412, 2021.

[15] O. Sami, Y. Elsheikh and F. Almasalha, "The role of data pre-processing techniques in improving machine learning accuracy for predicting coronary heart disease," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 812–820, 2021.

[16] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. J. Schmid *et al.,* "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989.

[17] J. Li, H. Lei, A. H. Alavi and G. G. Wang, "Elephant herding optimization: Variants, hybrids, and applications," *Mathematics*, vol. 8, no. 9, pp. 1415, 2020.

[18] R. Duwairi and A. Melhem, "A deep learning-based framework for automatic detection of drug resistance in tuberculosis patients," *Egyptian Informatics Journal*, vol. 24, no. 1, pp. 139–148, 2023.

[19] S. Murugesan, R. S. Bhuvaneswaran, H. Khanna Nehemiah, S. Keerthana Sankari and Y. Nancy Jane, "Feature selection and classification of clinical datasets using bioinspired algorithms and super learner," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–18, 2021.

[20] C. H. Lin, P. K. Yang, Y. C. Lin and P. K. Fu, "On machine learning models for heart disease diagnosis," in *Proc. of the 2nd Eurasia Conf. on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, Tainan, Taiwan, IEEE, pp. 158–161, 2020.

[21] M. J. A. Junaid and R. Kumar, "Data science and its application in heart disease prediction," in *Proc. of the Int. Conf. on Intelligent Engineering and Management (ICIEM)*, London, UK, IEEE, pp. 396–400, 2020.

[22] L. Ali and S. A. C. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *IRBM*, vol. 42, no. 5, pp. 345–352, 2021.

[23] S. Sandhiya and U. Palani, "An effective disease prediction system using incremental feature selection and temporal convolutional neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 5547–5560, 2020.

[24] S. P. Patro, G. S. Nayak and N. Padhy, "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective," *Informatics in Medicine Unlocked*, vol. 26, pp. 100696, 2021.

[25] M. A. Khan, "An IoT framework for heart disease prediction based on MDCNN classifier," *IEEE Access*, vol. 8, pp. 34717–34727, 2020.

[26] A. Al Bataineh and S. Manacek, "MLP-PSO hybrid algorithm for heart disease prediction," *Journal of Personalized Medicine*, vol. 12, no. 8, pp. 1208, 2022.

[27] A. Sharma, R. S. Tanwar, Y. Singh, A. Sharma, S. Daudra *et al.,* "Heart rate and blood pressure measurement based on photoplethysmogram signal using fast Fourier transform," *Computers and Electrical Engineering*, vol. 101, pp. 108057, 2022.

[28] S. L. Pandey, A. Shukla, S. Bhatia, T. R. Gadekallu, A. Kumar *et al.,* "Detection of arrhythmia heartbeats from ECG signal using wavelet transform-based CNN model," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, pp. 80, 2023.