



Optimized Hidden Count Deep Belief Network for Heart Disease Prediction using Hybrid Lion plus Firefly Algorithm

Harshith Chowdhary

Lamar University
Beaumont, Texas, United States.
iharshith512@gmail.com

Abstract: Due to the increasing count of healthcare data, the processing and analysis becomes more complicated particularly when comes to the maintenance of e-healthcare data. Under this circumstance, the disease prediction seems to be quite intricate in the medical application. This paper aims to introduce a new heart disease prediction model with phases like (i) Pre-processing, (ii) Proposed Feature Extraction, (iii) Dimensionality Reduction (iv) Classification as well. The first phase is pre-processing, where the data normalization process is carried out. From the pre-processed data, higher-order differential features get extracted by defining a new differential evaluation. The extracted features suffer from the 'Curse of Dimensionality', and are rectified by using Principle Component Analysis (PCA) that makes the dimensionality reduction proper. The dimensional reduced features are subjected to a classification process, where a deep learning model is used. To make the prediction model more accurate, it is planned to insist on the optimization logic in this, and thereby Optimized Deep Belief Network (DBN) model is used for prediction. Particularly, the count of hidden neurons in DBN is optimally selected by a new hybrid algorithm namely Firefly-based Improved Nomad Lion (FF-INL) that hybrids the concept of both the Lion Algorithm (LA) and Firefly (FF) algorithm. The simulation outcomes have thus revealed the efficacy of this proposed prediction work.

Keywords: Heart Disease; Pre-processing; Feature Extraction; Classification; Optimization

Nomenclature

Abbreviations	Descriptions
EMR	Electronic Medical Record
DE	Differential Evolution
AHP	Analytical Hierarchy Process
SVM	Support Vector Machines
IT2FLS	Interval Type-2 Fuzzy Logic System
k-NN	K Nearest Neighbor
LR	Logistic Regression
GA	Genetic Algorithm
NN	Neural Network
LAD	Left Anterior Descending
LCX	Left Circumflex Artery
RCA	Right Coronary Artery
DT	Decision Tree
IHD	Ischemic Heart Disease
LA	Lion Algorithm
FF	Firefly Algorithm
CD	Contrastive Divergence
Bi-LSTM	Bidirectional long short-term memory
PSO	Particle Swarm Optimization
MLP	Multilayer Perceptron
AL	Active Learning

1. Introduction

Because of the escalating growth in the healthcare data quantity [9] [10] [11] [12], valuable information is extracted by data analytics within a healthcare information system. At present, most healthcare organizations have established digitization services because of the enormous volume of healthcare data, which in turn leverages data analytics in healthcare to realize the massive benefits. The possible

advantage involves improved patient experience, early detection of diseases, personalized treatment, cost reduction, and superior care. A medical record is termed as a file that comprises information about patient characteristics and examinations, therapies, treatments, and services. The outcomes of the medical record's digitized translation are referred to as EMRs [13] [14]. These EMRs are complicated and have large sizes. Hence, the management of these large data volumes is very critical with the existing data processing models.

Heart disease has risen considerably over the past few decades and has become a huge threat to most people all over the world. Many characteristics exist in heart disease that cause an impact on the functioning of the heart. Therefore, it may be critical for doctors to cure that accurately and rapidly. So, computerized technologies in heart disease diagnosis [24] [25] is very much essential in healthcare terms for assisting the doctor to cure quicker with greater accuracy. The coronary disease examination is made on the perception of manifestations, signs, and physical examination of the patient. The most challenging and crucial tactic in restorative administrations is the determination of the correct illness. The prediction of coronary illness using medical parameters of patient affectionate tests is considered a multi-layered problem that can cause surprising effects and false suppositions.

Nowadays, more such heart disease diagnosis methods based on soft computing approaches have been introduced in the literature. But all are based on single-operation models. Hence, to perform superior operation over the single techniques in research work, more models incorporating the utilization of various soft computing approaches have been examined. Thereby, new hybrid models will be evolved. Generally, two stages consist of these models. The initial stage is the selection of subset features by applying the feature selection approaches [15] [16] [17] [18] [19]. The second stage is the classification process [20] [21] [22] [23], where algorithms like SVM, NN, and Naive Bayes, etc are used. However, the attainment of precise results is still a barrier in this field.

The contribution of this proposed model is explained as follows:

- In this research work, a new heart disease prediction model is implemented with four phases Pre-processing, Proposed Feature Extraction, Dimensionality Reduction, and Classification.
- The pre-processing is the initial phase, where the data normalization process is performed.
- After that, the higher-order differential features get extracted from the pre-processed data by explaining a new differential evaluation.
- The features thus extracted suffer from the 'Curse of Dimensionality', and this can be solved by employing PCA.
- The dimensional reduced features are then subjected to the classification process, where a deep learning model is deployed.
- Then the Optimized DBN model is used for prediction, where the count of hidden neurons in DBN is optimally selected by a new hybrid algorithm namely FF-INL that hybrids the theory of LA and FF algorithm.
- The evaluation outcome thus explains the efficacy of this proposed work.

The remaining paper is in this order as follows: Section 2 defines the literature review on the corresponding papers related to heart disease prediction. Section 3 depicts the proposed perceptive on the heart disease prediction model: proposed feature extraction process. Section 4 delineates the dimensionality reduction and optimized DBN-based classification. Section 5 explains the proposed hybrid model for optimal tuning of hidden neuron count hybridizing LA and FF. Section 6 expresses the results and their discussions. section 7 concludes the paper.

2. Literature Survey

2.1 Related Works

In 2024, S. Venkatesh Babu *et al.* [8] discussed the effectiveness of Quantum-Enhanced Machine Learning (QuEML) in predicting heart disease. It compared the performance of QuEML against traditional machine learning algorithms using the Kaggle heart disease dataset. The study highlighted the integration of quantum computing to enhance machine learning algorithms for faster computation in diagnosing heart diseases. Quantum computing principles like qubits, superposition, entanglement, quantum tunneling, and quantum gates are utilized to improve accuracy and computation time in disease prediction.

In 2024, Atta Ur Rahman *et al.* [1] used an enhanced heart disease prediction using a self-attention-based transformer model that combines self-attention mechanisms and transformer networks to predict cardiovascular disease risk. The model aimed to provide early diagnoses for individuals prone to heart failure by analyzing their characteristics and achieved a high accuracy of 96.51% on the Cleveland

dataset. The self-attention layers capture contextual information, generate representations to model complex patterns in the data, and provide interpretability by assigning attention weights to input components. The study also highlights the limitations of previous heart disease prediction methods, motivates the development of an improved model, and emphasizes the importance of effectively processing sequential data in healthcare applications.

In 2022, Raniya R. Sarraet *et al.* [2] executed based on the SVM algorithm. This algorithm was used for classification and regression tasks to detect heart disease. To improve the model's prediction accuracy, the two statistically optimal feature selection techniques were used. This technique was used to identify the highly related features in a dataset for a given task. The output of the method was then validated by comparing it to traditional models using several performance measures. For this model, they obtained more accurate results, and the component load was reduced by half.

In 2022, Angel Nancy *et al.* [3] systematically used predictive analytics to obtain information from clinical information that was stored electrically in the cloud and the Internet of Things equipment. A Bi-LSTM was assembled for monitoring and to accurately predict the risk of heart disease. This was a type of deep learning algorithm that was used to analyze sequential time-series data. It was able to process data quickly and accurately and can be used to make predictions about future outcomes. Bi-LSTM provided accurate predictions compared with the existing heart disease prediction systems.

In 2022, Nurul Absaret *et al.* [4] used machine learning algorithms to detect diseases related to the heart. It discussed four different models, which were tested on datasets from Kaggle. In the end, a cloud hosting platform using the internet, known as Stream lit was developed for predicting diseases. This plays a significant process in identifying diseases in the heart easily. This model contributed significantly to predictors in the prognosis of head diseases.

In 2022, Ali Al Bataineh and Sarah Manacek [5] experimented with different intelligent systems integrated with ML algorithms to identify diseases in the heart using a dataset. Initially, a PSO algorithm used techniques such as alternative MLP to detect heart disease. Here, ten different ML algorithms and an MLP-PSO hybrid algorithm were used for the prediction of heart diseases. The result was more precise and effective.

In 2022, Mohamed G. El-Shafieyet *et al.* [6] introduced machine-learning algorithms to detect heart diseases in the easy stage. Initially, PSO and GA were enhanced methods that were called GAPSO-RF. The GAPSO-RF was based on RF. Then multivariate statistical analysis was used for the initial population which was the first step. In the second step, GA implemented a discriminate mutation technique. In this case, GAPSO-RF combined local search with PSO and modified GA for global search to search the entire search space. Then, for rehabbing, PSO was used for the selection process to give candidates a second chance. Finally, the performance was validated.

In 2022, Ibrahim M. El-Hasnonyet *et al.* [7] included MMC, AUDI, AL, Random, Adaptive, and QUIRE were the first five multi-label active learning strategies to lower the costs of Lebling by repeating the greatest pertinent data for the next labels. Next, the grid-search optimization was done on the selected hyperparameters through a label ranking classifier. The experimental evaluation used was done through optimized settings and F-score was highlighted accurately without hyperparameter optimization.

2.2 Review

Table I represents the features and challenges of the conventional models regarding heart disease prediction. various techniques have been evolved in the literature for heart disease predictions like machine learning algorithms, data mining techniques, etc. still some issues exist in the literature that need to be rectified. The optimal criterion feature selection model [1] has reduced prediction time and eliminates irrelevant attributes. But, does not guarantee any improved performance. Modified DE algorithm [2] poses reduced prediction time, minimizes the critical attributes, and improves prediction accuracy. However, error minimization is still processed and the heart disease prediction is made in the small dataset. GA [3] generates an accurate rule set for heart disease risk level prediction and has better accuracy. Still, designing the objective function is critical and is computationally expensive. SVM [4] improves the prediction accuracy and identifies the patient at higher risk. Even though poses several key parameters and needs significant improvement in classification results. Chaos FF algorithm [5] has better convergence speed and processing time and has high classification accuracy. Still has some disadvantages like training time is quite slow and needs to achieve significant speedup. Mediative fuzzy logic [6] has the possibility of handling contradictory and doubtful information and the possibility of implementing a more realistic fuzzy interference system. Yet, needs improvement in the reasoning system and better tuning of the membership function. Naive Bayes [7] has higher prediction accuracy and better performance. But, still needs experimentation in large-scale real-world datasets. SVM [8] poses better processing of the whole QRS segment and improves the localization accuracy. Two major drawbacks are the choice of kernel is difficult and the extensive memory requirement.

Table 1. Features and Challenges of State-of-the-art Model on Heart Disease Prediction

Author [citation]	Methodology	Features	Challenges
S. Venkatesh Babu et al. [8]	Quantum-Enhanced Machine Learning	* Outperforms traditional machine learning approaches in terms of accuracy and computational time in heart disease prediction.	* lack support for better patient outcomes, indicating a need for improved heart disease prediction methods.
Atta Ur Rahman et al. [1]	self-attention-based transformer model	* Enhancing interpretability. * Effectively captures dependencies, enhances interpretability, and enables computation parallelization, reducing limitations of previous methods.	* Understanding the transformer's architecture, especially in deeper and more complex models, can be challenging for users.
Raniya R. Sarraet al. [2]	Support vector machine (SVM) algorithm	* Improved accuracy * Reduced componential load. * Better prognosis	* Limited dataset * Lack of interpretability * Need for further validation
Angel Nancy et al. [3]	IoT and DL	* Improved accuracy * Proactive healthcare strategy * Efficient data management * Real-time monitoring	* Data privacy concerns * Limited generalizability * Lack of real-world testing
Nurul Absaret al. [4]	Machine learning	* High accuracy * Low cost * Convenient and fast	* Limited datasets * Lack of diversity * Lack of explanation
Ali Al Bataineh and Sarah Manacek [5]	MLP-PSO hybrid algorithm and ML algorithms	* High accuracy and best performance.	* Does not discuss generalizability.
Mohamed G. El-Shafiey et al. [6]	hybrid genetic algorithm (GA) and particle swarm optimization (PSO)	* Achieved better accuracy than other state-of-the-art methods.	* Evaluated only on two public datasets - Cleveland and Statlog. * May not be suitable for all types of data sets or classification tasks.
Ibrahim M. El-Hasnonyet al. [7]	ML, Supervised learning, Active Learning, selection strategies, and Label ranking classifier	* Reduced labelling cost. * Improved classification quality. * Generalization beyond existing data.	* Lack of comparison with other models. * Limited dataset.

3. Proposed Perspective on Heart Disease Prediction Model: Proposed Feature Extraction Process

In this research work, a disease prediction approach is introduced for the effective prediction of heart problems in the patient. This prediction model is of four stages called, (i) Pre-processing (ii) Proposed Feature Extraction (iii) Dimensionality Reduction, and (iv) Classification. The healthcare data of heart disease patients are first subjected to pre-processing, where the normalization of data is exploited. The gathered data $D = \{d_1, d_2, \dots, d_m\}$ are normalized under the range 0 to 1. This normalization is

$$D_{M \times N} = \begin{Bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{Bmatrix}$$

explained as follows:

Normalization: Database Normalization is assumed as an approach to arrange the data within the database. This normalization process is made over the data, previous to the feature extraction process, and is defined using Eq. (1).

$$N = \frac{\hat{d}_{ij}}{\max(\hat{d}_{ij})} \quad (1)$$

$i=1 \text{ to } M, j=1 \text{ to } N$

Finally, the normalized data is obtained from the pre-processing phase. From this data, the proposed higher-order differential features are extracted in the feature extraction phase, and the description is explained in the following section. The extracted features are subjected to dimensionality reduction,

where PCA is deployed. After this, the dimensionality-reduced data is applied to the classification phase, where the data classification takes place using an optimized DBN model. As per the prime contribution, this paper intends to enhance the prediction model by using the optimization theory. Here, the count of hidden neurons in DBN is optimally selected by introducing a new algorithm named FF-INL, which is the hybridization of LA and FF. Fig. 1 shows the proposed heart disease, prediction model.

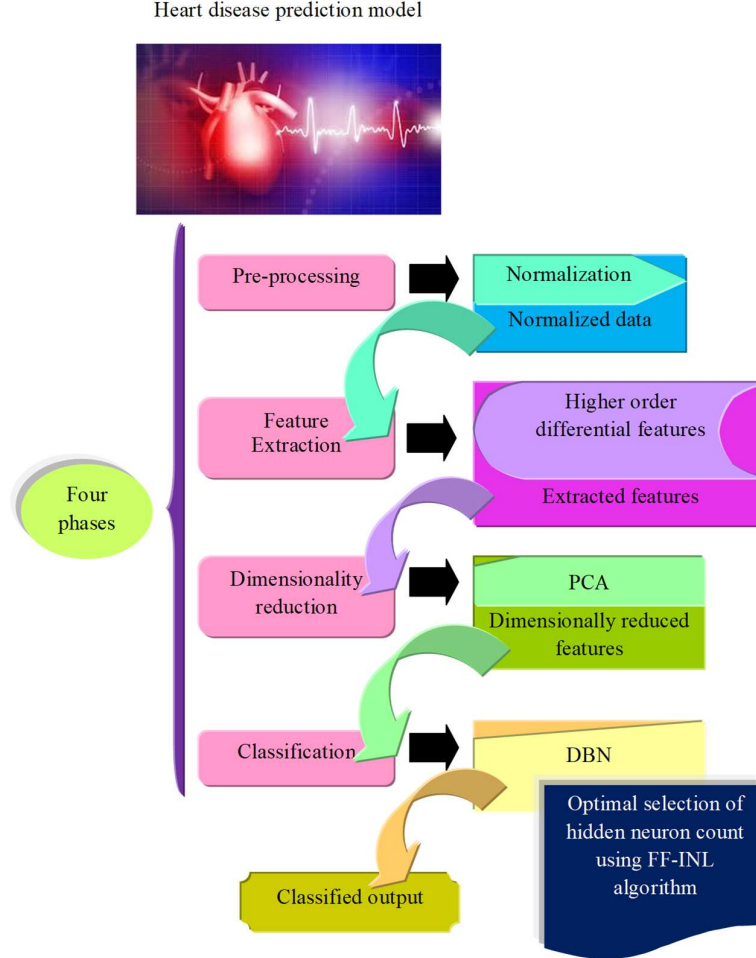


Fig. 1. Art ofproposed Heart Disease Prediction Model

3.1 Proposed Higher Oder Differential Feature Extraction

The main contribution of this paper relies on the feature extraction phase. In this, from the normalized data after pre-processing, the higher-order differential features are extracted. Eq. (2) depicts the higher-order differential equation, where $x_{ij} : i = 1, 2, \dots, N_R$ denotes the number of records and $x_{ij} : j = 1, 2, \dots, N_A$ indicates the count of attributes.

$$\mu x_{ij} = \frac{1}{N_A - 1} \sum_{k=1, k \neq j}^{N_A} (x_{ij} - x_{ik}) \quad (2)$$

From this, the first-order differential features are extracted and are expressed as per Eq. (3). Similarly, the second and third-order differential features $\delta x_{ij}^{(2)}$ and $\delta x_{ij}^{(3)}$ are also extracted as per the evaluation given in Eq. (3).

$$\delta x_{ij}^{(1)} = \mu x_{ij} \quad (3)$$

The pseudo-code of N-order differential feature extraction is explained in Algorithm 1.

Algorithm 1: N-order differential feature extraction	
i/p: X /o/p: $\hat{X}$	
$\mu x_{ij}^{(0)} \leftarrow x_{ij}$	
For $n = 1$ to N	
Calculate the distance matrix for $\mu x_{ij}^{(n)}$ as $D_i = d_{jk}^{(i)}$ such that $\ d_{jk}^{(i)}\ = N_A \times N_A$	
Calculate the mean difference vector $\mu x_{ij}^{(n)}$ as $\delta x_{ij}^{(1)} \leftarrow \mu x_{ij}^{(n)}$	
End for	
$\hat{X} = [X \ \delta X]$ such that $\ \delta X\ = N_R \times N_D$	

3.2 Relationship Model Between Conventional and Proposed Features

This section explains the correlation between conventional and proposed features by the graphical representation in Fig 2.

The evaluation is as per the following process.

- The actual features are summed up with the features obtained using Eq. (3), together form the first-order features.
- These first-order features are initially correlated and averaged to get one average count.
- Each actual and first-order features are compared to the average count, and the features that exceed the average count are plotted in the graph given in Fig. 2.
- From the graph, it is proved that the proposed first-order differential features have attained maximum correlation and correlation range than the existing features.

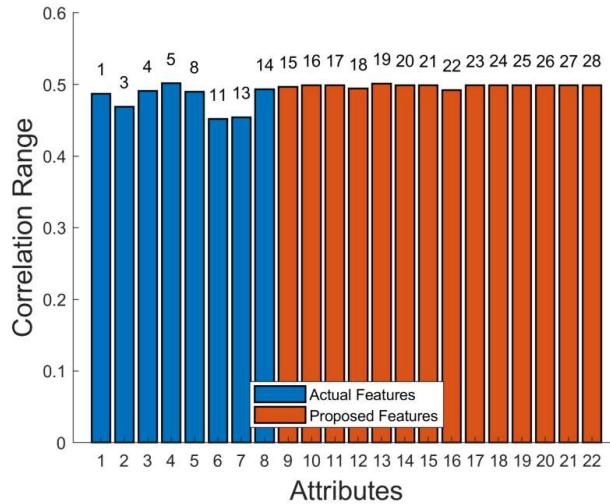


Fig. 2. Relationship between the Actual Features and Proposed First order Features

4. Dimensionality Reduction and Optimized DBN-based Classification

4.1 PCA for Dimensionality Reduction

Once the feature extraction phase was completed, the extracted higher-order statistical features underwent dimensionality reduction using PCA due to the vast quantity of features. The PCA is the common name for an approach that utilizes sophisticated basic arithmetical principles for transforming a count of probably correlated variables into a smaller count of variables named principal components [8]. The main intention of PCA is to minimize the data space's large dimensionality into a lesser feature space's intrinsic dimensionality, which in turn requires explaining the data economically. The arithmetical formulation of PCA is exemplified by the following:

Mean: Assume $Y_1, Y_2, \dots, Y_n$ depicts the arbitrary variables for sample size n . The data set mean is an arbitrary variable and is explained as per Eq. (4).

$$Mean(\bar{X}) = \frac{1}{n} \sum_{i=1}^n Y_i \quad (4)$$

Standard Deviation: STD dev is the measure of the data spread out. The average distance from the dataset mean to the point is needed to calculate the STD dev, i.e. estimate the distance square from every data point to the mean of the set, summarize the entire data, divide by n and use the positive square root, which is expressed in Eq. (5).

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (5)$$

Covariance: Covariance is the measurement between two dimensions. The covariance formula is somewhat the same as the variance and is stated in Eq. (6).

$$Cov(Y, G) = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(G_i - \bar{G})}{n} \quad (6)$$

Eigenvalues and Eigenvectors: If G is considered as a $n \times n$ matrix, then $[Y] \neq \bar{0}$ is an eigenvector of G is estimated as per Eq. (7), where λ is considered as scalar and $[Y] \neq 0$. The scalar λ is termed as an eigenvalue of G and $[Y]$ is referred to as the eigenvector according to eigenvalue λ . The eigenvalues of a $n \times n$ matrix G having a non-zero solution are stated in Eq. (8).

$$[G][Y] = \lambda[Y] \quad (7)$$

$$\det([G] - \lambda I) = 0 \quad (8)$$

The obtained output from the PCA is the dimensionality-reduced features.

4.2 Optimized DBN for Classification

The dimensionality-reduced data is then subjected to the classification process, where the DBN model is used for this purpose. DBN [21] [22] is considered a fascinating model that involves multiple layers with visible and hidden neurons in each layer. Further, the entire hidden and input neurons are connected, still, no connection persists within hidden neurons as well as visible neurons. In the Boltzmann network, since the stochastic neuron's output is probabilistic, Eq. (9) denotes the output, and Eq. (10) explains the probability within the sigmoid-shaped function, in which $Temp$ symbolizes the pseudo-temperature. The deterministic form of the stochastic model is delineated using Eq. (11).

$$Pb(\zeta) = \frac{1}{1 + e^{\frac{-\zeta}{Temp}}} \quad (9)$$

$$F = \begin{cases} 1 & \text{with } 1 - Pb(\zeta) \\ 0 & \text{with } Pb(\zeta) \end{cases} \quad (10)$$

$$\lim_{Temp \rightarrow 0^+} Pb(\zeta) = \lim_{Temp \rightarrow 0^+} \frac{1}{1 + e^{\frac{-\zeta}{Temp}}} = \begin{cases} 0 & \text{for } \zeta < 0 \\ \frac{1}{2} & \text{for } \zeta = 0 \\ 1 & \text{for } \zeta > 0 \end{cases} \quad (11)$$

Fig. 2 explains the DBN model. Eq. (12) explains the arithmetical approach to exposing the Boltzmann machine's energy for the composition of neuron state (binary state) c , here $wh_{i,j}$ explains the weights amongst the neurons and θ_i demonstrates the biases.

$$\Delta E(c_i) = \sum_j c_i wh_{i,j} + \theta_i \quad (12)$$

The energy definitions in the view of joint work of visible and hidden neurons (m, n) are mentioned in Eq. (13), Eq. (14) and Eq. (15), where, m_i exemplifies the binary state of the visible unit $\hat{i}$, h_j depicts the binary state of hidden unit $\hat{j}$ and g_i and h_i portrays the biases adopted over the network. In fact, the optimization concept is insisted in the tuning of count of hidden neurons. Rather than using the default count, the optimal selection makes the model more effective with respect to time consumption and preciseness. For this optimal tuning, a new hybrid algorithm is introduced in this paper, and it is explained in the upcoming section.

$$E(m, n) = \sum_{(i,j)} wh_{i,j} m_i n_j - \sum_i g_i m_i - \sum_j h_j n_j \quad (13)$$

$$\Delta E(m_i, \vec{n}) = \sum_j wh_{ij} n_j + g_i \quad (14)$$

$$\Delta E(\vec{m}, n_i) = \sum_j wh_{ij} m_j + h_i \quad (15)$$

The input data's probability distribution is programmed as weight parameters and is allotted as the learning pattern of RBM. Apparently, RBM training can increase the allotted probabilities, and the assignment on weight is estimated based on Eq. (16).

$$Wh^{(a)} = \max_{Wh} \prod_{\vec{m} \in A} p(\vec{m}) \quad (16)$$

For each practicable visible and hidden vector's pairs, the probability allotted RBM method is stated as per Eq. (17), where F denotes the partition function given in Eq. (18).

$$p(\vec{m}, \vec{n}) = \frac{1}{F} e^{-E(\vec{m}, \vec{n})} \quad (17)$$

$$F = \sum_{\vec{m}, \vec{n}} e^{-E(\vec{m}, \vec{n})} \quad (18)$$

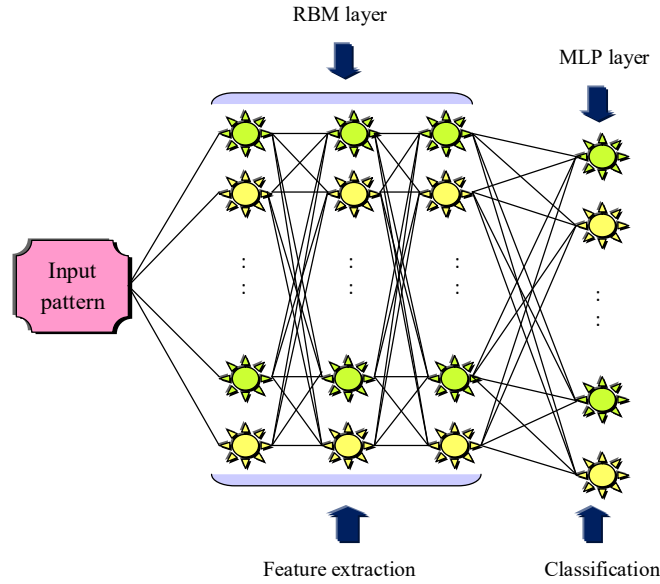


Fig. 3. The architecture of the DBN model

As it is a complicated to gain the expected sampling under the distribution explained by the format, CD learning method is employed. Owing to this, the step by step description of CD algorithm is as follows.

1. Choose on the training samples m and brace it within the visible neurons.
2. Evaluate the hidden neuron's probability p_n by determining the weight matrix Wh and visible vector m product by $p_n = \sigma(mWh)$, as per Eq. (19).

$$p(\vec{n}_j \rightarrow 1 | \vec{m}) = \sigma\left(h_j + \sum_i m_i wh_{i,j}\right) \quad (19)$$

3. Validate the hidden states n from the probabilities p_n .
4. Evaluate the exterior product of vectors m and p_n , state that as positive gradient $\phi^+ = m \cdot p_n^{Tp}$.
5. Validate the visible state's reconstruction m' from the hidden states n using Eq. (20). Moreover, the testing of hidden states n' from the visible state's reconstruction m' is needed.

$$p(\vec{m}_j \rightarrow 1 | \vec{n}) = \sigma\left(g_j + \sum_i n_i wh_{i,j}\right) \quad (20)$$

6. Estimate the exterior product of m' and n' , state that as negative gradient $\phi^- = m' \cdot n'^{Tp}$.
7. Find out the updated weight as per the Eq. (21), where η depicts the learning rate.

$$\Delta Wh = \eta(\phi^+ - \phi^-) \quad (21)$$

8. Update the weights with new values using Eq. (22).

$$wh'_{i,j} = \Delta wh_{i,j} + wh_{i,j} \quad (22)$$

For beginning the learning process using MLP approach, consider the training patterns (A^z, B^z) , where z exemplifies the count of training patterns, $1 \leq z < Z$, A^z and B^z delineates the vector input and expected vector output, respectively. The error of every neuron in $\hat{j}$ within the output layer is evaluated using Eq. (23).

$$er_j^z = A^z - B^z \quad (23)$$

Thus, Eq. (24) offers the pattern's squared error z sustained the mean squared error as per Eq. (25).

$$Er_z^{mean} = \frac{1}{nc_o} \sum_{j=1}^{nc_o} (er_j^z)^2 = \frac{1}{nc_o} \sum_{j=1}^{nc_o} (A^z - B^z)^2 \quad (24)$$

$$Er_{avg} = \frac{1}{Z} Er_z^{mean} \quad (25)$$

The DBN training process with pre-training (RBM) and normal training (MLP) is explained below.

1. Based on the arbitrarily selected weights, biases and other appropriate parameters, the DBN model can be initialized.
2. The initiation of RBM model is performed using the input data, serving potentials within its visible neurons and takes on the unsupervised learning.
3. The next layer input is gained by sampling the potentials presented in the hidden neurons of the earlier layer. Subsequently, it pursues the unsupervised learning.
4. These steps can be iterated for an exact count of layers. Thereby the pre-training stage by RBM is completed until it achieves the MLP layer.
5. MLP phase has offered refined learning using supervised format and is continued till the obtainment of target error rate.

5. Proposed Hybrid Model for Optimal Tuning of Hidden Neuron Count: Hybridizing LA and FF

5.1 Solution Encoding and Objective Function

This paper makes the contribution on the optimal tuning of hidden neurons in DBN. To make this possible, a new hybrid algorithm named, FF-INL is implanted in this work. The input solution to the proposed FF-INL is shown in Fig 4. Further, the defined fitness or objective of this work is given in Eq. (26), where A^p indicates the prediction accuracy.

$$obj = \max(A^p) \quad (26)$$

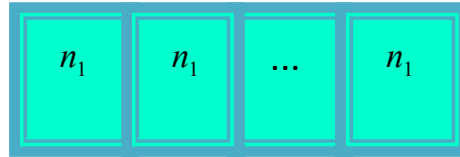


Fig. 4. Solution Encoding

5.2 Traditional LA Algorithm

LA is the nature-inspired algorithm that introduced based on the motivation by the social behavior of Lions. It is known that the Lions are survived as groups having one male head as pride. Male cubs join the birth pride still it became adult and then it leaves the pride and wander around as nomadic lions. On this wandering stage, if the nomadic lion supposes to meet the pride, then the fight will incur for supremacy. At the time of this encounter, the nomadic lion is suspected as pride, if it wins the pride. Every lion specifies each solution within the LA algorithm. The four main stages of LA algorithm are mating, pride generation, territorial takeover, and territorial defense.

Problem formulation: The intention of this LA model is assumed in Eq. (27). The continuous unimodal or multimodal function is explained as $f(\bullet)$ whereby the solution space size is given as $\Re^k$, here the real numbers are exploited using $\Re$ and $p_i : i = 1, 2, \dots, k$ is defined as the solution variable at i^{th} position and the solution vector's dimension is denoted as k , the i^{th} position variable's minimum and maximum limits are explained as $p_i^{\min}$ and $p_i^{\max}$, correspondingly. Eq. (28) indicates the optimal solution, which is given as P^{opt} . The solution space size is exploited by means of Eq. (29).

$$P^{opt} = \arg \min_{p_i \in (p_i^{\min}, p_i^{\max})} f(p_1, p_2, \dots, p_k); \quad k \geq 1 \quad (27)$$

$$\mathfrak{R}^P = \prod_{i=1}^p (p_i^{\min} - p_i^{\max}) \quad (28)$$

$$P^{opt} = P : f(P) < f(P' | P' \neq P; p'_i \in (p_i^{\min}, p_i^{\max})) \quad (29)$$

In this, the solution vector is depicted by P and is demonstrated as $P = [p_1, p_2, \dots, p_k]$.

Pride generation: The pride initiation is exploited regarding the territorial lion P^{ma} and lioness P^{fe} and a nomadic lion P^{nd} . The pride generation process takes the nomadic lion into consideration inspite of its non-membership. The depiction of lion is the same as that of solution vector. The vector element of P^{ma} , P^{fe} and P^{nd} is formulated as p_{ll}^{ma} , p_{ll}^{fe} and p_{ll}^{nd} area and is considered as the arbitrary integers that pose minimum and maximum limits while $k > 1$, here $ll = 1, 2, \dots, L$, where L exemplifies the length of the lion as per Eq. (30). k and q are integers. When $k = 1$ the algorithm is expected using Eq. (31) and (32).

$$L = \begin{cases} k; & k > 1 \\ q; & otherwise \end{cases} \quad (30)$$

$$l(p_{ll}) \in (p_{ll}^{\min}, p_{ll}^{\max}) \quad (31)$$

$$q \% 2 = 0 \quad (32)$$

where,

$$l(p_{ll}) = \sum_{ll=1}^L p_{ll} 2^{\left(\frac{ll}{2} - ll\right)} \quad (33)$$

Fertility Evaluation: The local or global optima is achieved by the laggards when P^{ma} and P^{fe} becomes saturated by the fitness. In order to leave out from local optimal solutions, the fertility evaluation is utilized. In this, the laggard is driven as P^{ma} and the laggardness rate is depicted as Ld_f and is increased by 1, when $f(P^{ma})$ achieves greater value than f^{rfr} , which defines the reference fitness. The territorial defense can be performed when Ld_f goes above its maximum limit $Ld_f^{\max}$. Once after the completion of crossover, the fertility of P^{fe} is make certain using sterility rate Sl_f and is increased by 1. Once Sl_f goes above the tolerance $Sl_f^{\max}$, then the updating of P^{fe} is proliferated using Eq. (34). Since because of the enrichment, the mating process is performed, when the updated female P^{fe+} is taken as P^{fe} .

$$p_{ll}^{fe+} = \begin{cases} p_b^{fe+} & \text{if } ll > b \\ p_{lel}^{fe} & \text{otherwise} \end{cases} \quad (34)$$

$$p_b^{fe+} = \min[p_b^{\max}, \max(p_b^{\min}, \nabla_b)] \quad (35)$$

$$\nabla_b = [p_b^{fe} + (0.1ra_2 - 0.05)(p_b^{ma} - ra_1 p_b^{fe})] \quad (36)$$

Where, p_{ll}^{fe+} and p_b^{fe} are considered to be ll^{th} and b^{th} vector element of P^{fe+} , correspondingly. The random integer b is formulated within the intervals $[1, L]$, the update function of female is given by ∇ , ra_1 and ra_2 are the arbitrary integers that produced in the interval $[0, 1]$.

Mating: In the mating process, crossover and mutation are considered as the two main phases and thereby the cubpools are created by the parents P^{ma} and P^{fe} . Here, the crossover process produced four cubs and the rest four by mutation process.

Lion operators: In the territorial defense, the wide searching of solution space is facilitated along with the guidance to algorithm for escaping from local optimum point and also finds the different solutions with the same fitness. The territorial defense might employ with nomad coalition, pride and nomad coalition update and survival fight. Further, the selection of P^{e-nd} is formulated only if the strategy as per Eq. (37)-(39) are satisfied.

$$f(P^{e-nd}) < f(P^{ma}) \quad (37)$$

$$f(P^{e-nd}) < f(P^{ma-cub}) \quad (38)$$

$$f(P^{e-nd}) < f(P^{fe-cub}) \quad (39)$$

The pride gets updated when P^{ma} is conquered, likewise, nomad coalition gets updated when P^{e_nd} is conquered. Territorial takeover process executes the algorithm for updating P^{ma} and P^{fe} , when P^{ma_cub} and P^{fe_cub} are matured. This is explained as after the cub's age goes beyond the maximum age limit of cub maturity, which is denoted as $A_{\max}$.

Termination: if anyone equation among Eq. (40) and Eq. (41) are satisfied, the algorithm gets terminated.

$$No_{gen} > No_{gen}^{\max} \quad (40)$$

$$f(P^{ma}) - f(P^{opt}) \leq er_T \quad (41)$$

where, the generation count is depicted as No_{gen} , which is zero initially and after the territorial takeover process, this will be increased by 1. The maximum generation count and error threshold is denoted as $No_{gen}^{\max}$ and er_T , correspondingly. Algorithm 2 explains the pseudo-code of conventional LA algorithm.

Algorithm 2: Conventional Lion Algorithm
Initialize P^{me} , P^{fe} and P^{nd}
Calculate $f(P^{me})$, $f(P^{fe})$ and P_1^{nd}
Assign $f^{tr} = f(P^{me})$ and $No_{gen} = 0$
Store P^{me} and $f(P^{me})$
Carry out the fertility function
Execute mating and produce cubpool
Compute gender clustering and gain P^{me_cub} and P^{fe_cub}
Initialize $A_{\max}$ as zero
Evaluate cub growth generation
Execute the territorial defense when defense results to zero, go to step 4
When $A_{cub} < A_{\max}$ going to step 9
Evaluate territorial takeover and gain the updated A^{me} and A^{fe}
$No_{gen} = No_{gen} + 1$
Go to step 4, if the termination criteria not met
Terminate the process

5.3 Conventional FF Algorithm

FF algorithm is based on the motivation of fireflies' social behavior in tropical sky. The communication, prey searching, and determining mates of fireflies are performed using the bioluminescence having diverse emission patterns. The three sets of rules under fireflies are explained as follows:

- Fireflies are unisex that mate with each other despite sex.
- Attractiveness and brightness are directly proportional and are reduced as per the increase in the distance of fireflies.
- The brightness and cost functions are related somehow, whereas the brightness is directly proportional to the cost function value.

The fireflies' initial position within the search space is determined as per Eq. (42), here i^{th} variable for the j^{th} agent having the initial value is delineated as $p_{i,j}^{(o)}$, the minimum and maximum permitted values of i^{th} the variable are explained by $p_{i,\min}$ and $p_{i,\max}$ respectively.

$$p_{i,j}^{(o)} = p_{i,\min} + rand.(p_{i,\max} - p_{i,\min}), i = 1, 2, \dots, N \quad (42)$$

Two noteworthy issues exist in the FF algorithm: attractiveness formation and light intensity variation. The light intensity can get varied in accordance to Eq. (43), where, the light intensity is specified by It , the original light intensity is referred to as It_0 and the coefficient of light absorption, which is constant and is depicted by γ . The attractiveness of the firefly is evaluated using Eq. (44), where β_0 is constant at attractiveness $u = 0$. The Cartesian distance is termed as the distance between

any two fireflies i and j at p_i and p_j correspondingly. The i^{th} fireflies' movement toward other attractive fireflies j is calculated as per Eq. (45). The attraction is exemplified using the first term and randomization is explained in the second one, where α is the randomization parameter and the arbitrary number extracted from Gaussian distribution is stated as a vector ε_i . Moreover, a geometric annealing scheduling function is deployed from the initial α_0 , and that is estimated using Eq. (46), where the randomness reduction constant is depicted by $0 < \theta < 1$.

$$I_t(\vec{d}) = I_o e^{-\gamma d^2} \quad (43)$$

$$\beta(\vec{d}) = \beta_o e^{-\gamma d^2} \quad (44)$$

$$\Delta p_i = \beta_o e^{-\gamma d_{ij}^2} (p_j^t - p_i^t) + \alpha \varepsilon_i, p_i^{t+1} = p_i^t + \Delta p_i \quad (45)$$

$$\alpha = \alpha_0 \theta^t \quad (46)$$

5.4 Proposed FF-INL Algorithm

LA is the renowned bio-inspired optimization algorithm that was developed based on Lion's social behavior. This can effectively solve single-objective and multi-objective optimization problems, whereas the FF algorithm is also the nature-inspired algorithm based on the social behavior of Fireflies. Though both algorithm has much positive impact on many works, they still pose some issues like premature convergence and as well minimize the chance to determine the global optimum. Hence, to overcome all these issues that existed; this paper aims to propose a new hybrid optimization algorithm that incorporates the process of FF algorithm in LA algorithm and termed as FF-INL. The procedure of the proposed algorithm is as follows: The nomadic lion generation of LA gets changed with the FF process. Particularly, the traditional LA algorithm initializes P^{nd} randomly, whereas the proposed algorithm makes the initialization using the update evaluation of the FF algorithm given in Eq. (46). The pseudo-code and flowchart of the proposed FF-INL algorithm are represented in Algorithm 3 and Fig. 4, respectively.

Algorithm 3: Proposed FF-INL Algorithm
Initialize P^{me} and P^{fe}
Initialize the P^{nd} as per the firefly update in Eq. (46)
Calculate $f(P^{me})$ and $f(P^{fe})$
Generate P_1^{nd} as per the Firefly update in Eq. (46)
Assign $f^{rfr} = f(P^{me})$ and $No_{gen} = 0$
Store P^{me} and $f(P^{me})$
Carry out the fertility function
Execute mating and produce cubpool
Compute gender clustering and gain P^{me_cub} and P^{fe_cub}
Initialize A_{max} as zero
Evaluate cub growth generation
Execute the territorial defense when defense results to zero, go to step 4
When $A_{cub} < A_{max}$, go to step 9
Evaluate territorial takeover and gain the updated A^{me} and A^{fe}
$No_{gen} = No_{gen} + 1$
Go to step 4, if the termination criteria not met
Terminate the process

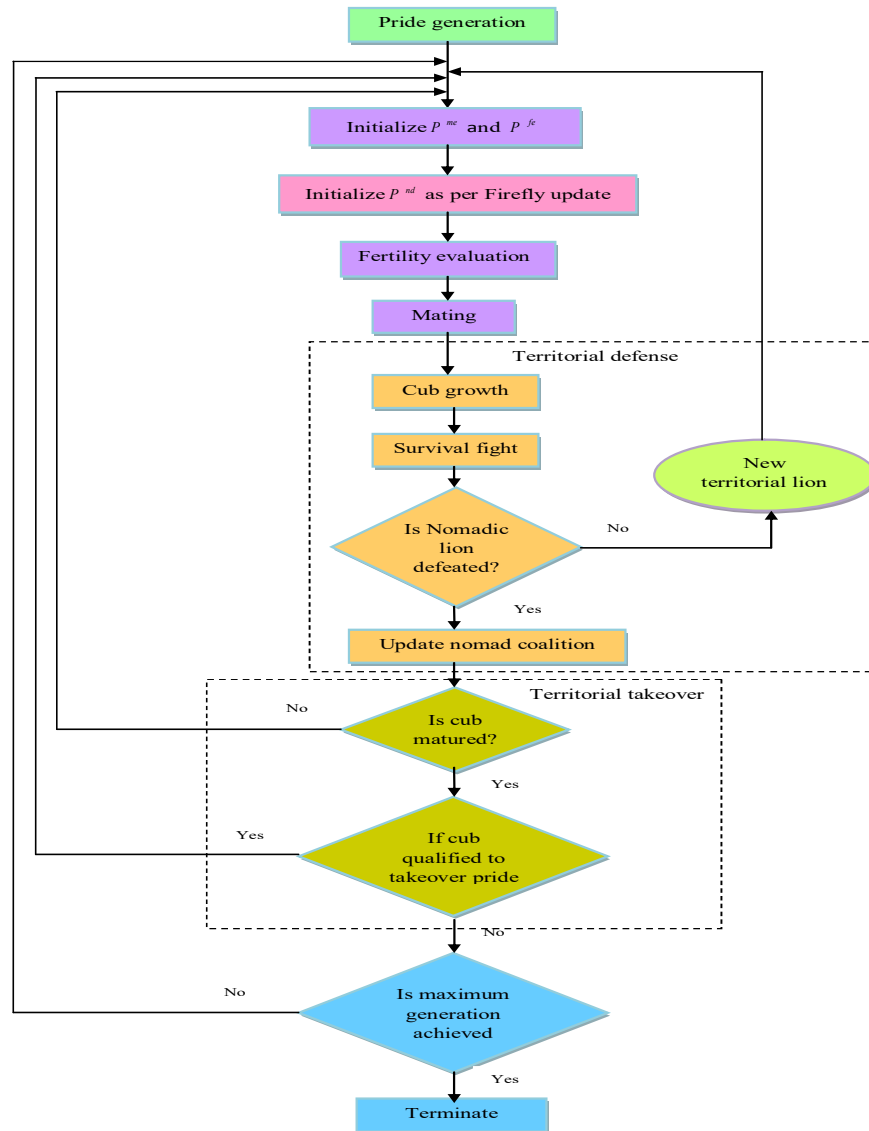


Fig. 5. Flowchart demonstration of Proposed FF-INL Algorithm

6. Results and Discussions

6.1 Simulation Setup

The simulation of this proposed heart disease prediction model was implemented using MATLAB. The feature set used in this work was downloaded from [http://archive.ics.uci.edu/ml/datasets/statlog+\(heart\)](http://archive.ics.uci.edu/ml/datasets/statlog+(heart)). Further, the analysis has been made on the performance in terms of the count of the attribute, hidden neurons, and features. Moreover, the comparative performance was handled among the actual features with first-order and second-order differential features. Finally, the overall performance has been analyzed regarding the positive measure like accuracy, sensitivity, specificity, precision, NPV, F1-Score, and MCC, as well in terms of negative measures like FPR, FDR, and FNR.

6.2 Performance Analysis Regarding Count of Attributes

Fig. 6 delineates the analysis of the performance measure proposed over conventional in the view of the count of attributes. In this, the performance of the proposed model is compared over the conventional model named MDE in [2] and the results are analyzed. In the existing work [2], the integrated MDE model using 9 critical attributes is compared to overall available input attributes. In this proposed work, the proposed 14 attributes are differentiated over the 9 critical attributes and 13 available attributes. In the conventional model, the overall accuracy obtained is 83%, whereas the proposed work has gained 95%, which is 14.46% better than the conventional model. Similarly, the other measures like sensitivity, specificity, and precision have achieved superior performance than those of conventional attributes and

are given by 96%, 96%, and 95.5% respectively. Thus from this analysis, it is clear that the proposed work outperforms the existing work by attaining precise accuracy.

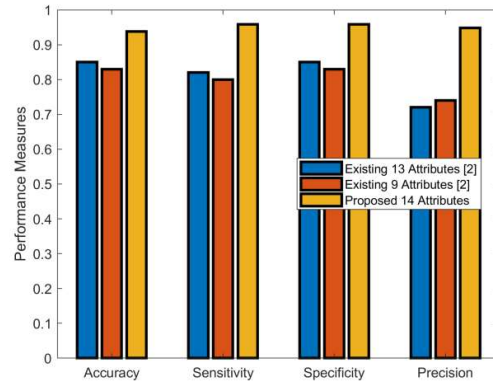


Fig. 6. Performance analysis of Proposed and Conventional models regarding the count of features

6.3 Actual Vs First-Order Features and Second-Order Features

Fig. 7 exemplifies the comparison between the actual features with the first-order and second-order features. The actual normalized features lie in the range 0 to 1, as from the graph, it can seem that the range among the actual features is high. After applying the first-order differential functions, the features are ranged over a low value and that is illustrated in Fig. 7(a). From the graph, it can say, the maximum range of features lies between 0.1 and 0.2. Further, after undergoing the second-order differential functions, the features ranged from 0.05 to 0.1, which is even lower than the first-order differential features, given in Fig. 7(b). This analysis thus reveals the better ranging of features, when going through higher-order differential features.

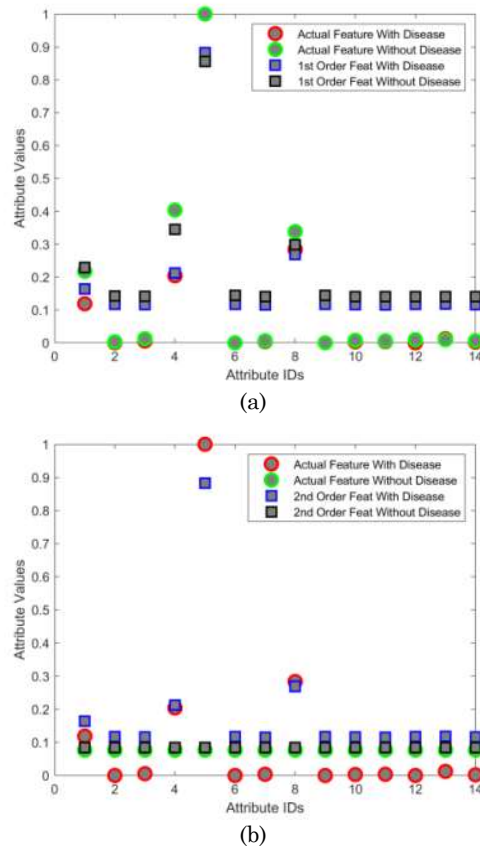


Fig. 7. Graph representation (a) Actual versus First order features (b) Actual versus Second-order features

6.4 Performance Analysis based on Hidden neurons

Fig. 9 depicts the analysis of the performance measure of the proposed model based on hidden neurons in DBN. Initially, the hidden neurons are assigned as a fixed value, i.e. 5 and 10 and the performance is evaluated. Similarly, another evaluation is performed by applying the optimal hidden neuron solution (Proposed), and the performance is evaluated. These two performances are compared together and the resultant outcome explains the superiority of the proposed work over the traditional models in terms of higher accuracy, sensitivity, specificity, and precision.

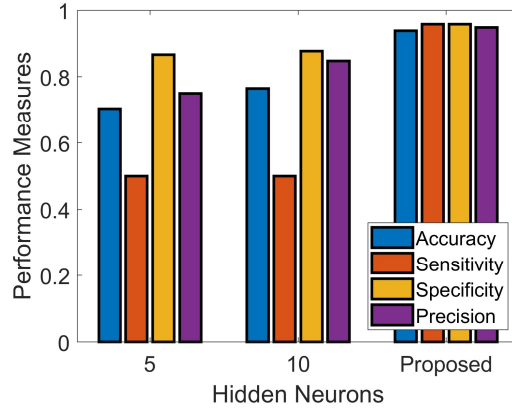


Fig. 8. Performance analysis of the proposed model with and without optimization

6.5 Performance Analysis in Terms of Features

Fig. 10 explains the performance analysis of the proposed model among the actual and higher-order differential features. In this, the actual features are compared with the first-order, second-order, and third-order differential features and are plotted in the graph respectively. When differentiated over the actual features, on every higher-order differential feature, the performance in terms of accuracy, sensitivity, specificity, and precision has gained higher performance than the previous higher-order features. From this, it is validated that the proposed model with higher-order differential features has shown better outcomes with higher accuracy.

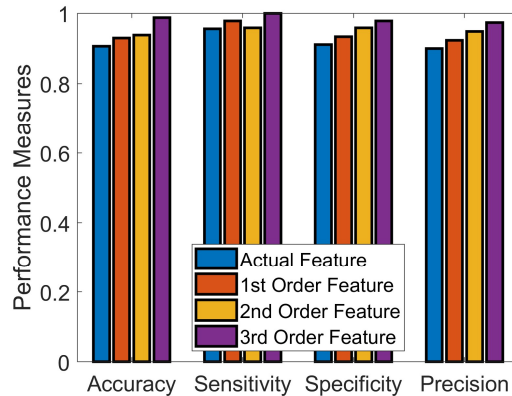


Fig. 9. Performance analysis of the proposed model between actual and higher-order differential features

6.6 Overall Performance Analysis

Table II specifies the overall analysis of the performance of the proposed model over the conventional model. In this, comparative performance is made through positive and negative measures. From the table, the performance of the introduced model has achieved improved accuracy, which is 15.15%, 13.43%, 10.14%, 16.92%, and 20.63% better than GA, PSO, WOA, LA, and FF respectively. For the FDR measure, the developed model is 75.35%, 74.8%, 69.33%, 78.09%, and 78.97% superior to GA, PSO, WOA, LA, and FF respectively. Owing to the MCC measure, the presented model has achieved better performance than GA, PSO, WOA, LA, and FF by 34.85%, 28.93%, 22.99%, 34.2%, and 48.59% respectively. From this analysis, it is clear that the proposed model outperforms the conventional models regarding the performance measure.

Table 2. Overall Performance analysis of the Proposed model over the conventional model

Measure	GA [28]	PSO [29]	WOA [30]	LA [26]	FF [27]	FF-INL
Accuracy	0.81481	0.82716	0.85185	0.80247	0.77778	0.93827
Sensitivity	0.91667	0.94444	0.91667	0.97222	0.91667	0.95827
Specificity	0.73333	0.73333	0.8	0.66667	0.66667	0.95827
Precision	0.73333	0.73913	0.78571	0.7	0.6875	0.94827
FPR	0.26667	0.26667	0.2	0.33333	0.33333	0.061728
FNR	0.083333	0.055556	0.083333	0.027778	0.083333	0.054728
NPV	0.73333	0.73333	0.8	0.66667	0.66667	0.93827
FDR	0.26667	0.26087	0.21429	0.3	0.3125	0.065728
F1_Score	0.81481	0.82927	0.84615	0.81395	0.78571	0.92727
MCC	0.65	0.67988	0.71272	0.65316	0.58992	0.87654

7. Conclusion

This framework is concerned with introducing a novel heart disease prediction model that includes four phases such as: (i) Pre-processing, (ii) Proposed Feature Extraction, (iii) Dimensionality Reduction, and (iv) Classification. The pre-processing was executed as the first phase, where the data normalization process was evaluated. After this pre-processing, the higher-order differential features were extracted based on the new differential evaluation. These extracted features suffered from the ‘Curse of Dimensionality’, and hence were solved by deploying the PCA method. The classification process was the subsequent step, where a deep learning model was utilized. Finally, the optimization concept was used for the precise prediction, in which the hidden neurons in DBN were optimally selected by intruding a new FF-INL algorithm, which combines the theory of LA and FF. From the simulation, it was observed that the overall accuracy obtained in the conventional model is 83%, whereas the proposed work has gained 95%, which is 14.46% better than the conventional model. Similarly, the other measures like sensitivity, specificity, and precision have achieved superior performance than those of conventional attributes and are given by 96%, 96%, and 95.5% respectively.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

- [1] Rahman, A.U., Alsenani, Y., Zafar, A., Ullah, K., Rabie, K., and Shongwe, T., “Enhancing heart disease prediction using a self-attention-based transformer model”, Scientific Reports, Vol. 14(1), p.514, 2024.
- [2] Cherian, R.P., Thomas, N., and Venkitachalam, S., “Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm”, Journal of Biomedical Informatics, Vol. 110, pp.103543, 2020.
- [3] Nancy, A.A., Ravindran, D., Raj Vincent, P.D., Srinivasan, K., and Gutierrez Reina, D., “Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics”, Vol. 11(15), pp.2292, 2022.
- [4] Absar, N., Das, E.K., Shoma, S.N., Khandaker, M.U., Miraz, M.H., Faruque, M.R.I., Tamam, N., Sulieman, A., and Pathan, R.K., “The efficacy of machine-learning-supported smart system for heart disease prediction”, In Healthcare (Vol. 10, No. 6, p. 1137). MDPI, 2022.
- [5] Al Bataineh, A. and Manacek, S., “MLP-PSO hybrid algorithm for heart disease prediction”, Journal of Personalized Medicine, Vol. 12(8), pp.1208, 2022.
- [6] El-Shafiey, M.G., Hagag, A., El-Dahshan, E.S.A. and Ismail, M.A., “A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest”, Multimedia Tools and Applications, Vol. 81(13), pp.18155-18179, 2022.
- [7] El-Hasnony, I.M., Elzeki, O.M., Alshehri, A., and Salem, H., “Multi-label active learning-based machine learning model for heart disease prediction”, Sensors, Vol. 22(3), pp.1184, 2022.
- [8] Babu, S.V., Ramya, P. and Gracewell, J., “Revolutionizing heart disease prediction with quantum-enhanced machine learning”, Scientific Reports, Vol. 14(1), pp.7453, 2024.
- [9] Claus, Bossen, Yunan, Chen, Kathleen H.Pine, "The emergence of new data work occupations in healthcare: The case of medical scribes", International Journal of Medical Informatics, vol.123, pp.76-83, March 2019.
- [10] Vasileios C. Pezoulas, Konstantina D.Kourou, Fanis Kalatzis, Themis P.Exarchos, Alik Venetsanopoulou, Evi Zampeli, Saviana Gandolfo, Fotini Skopouli, Salvatore De Vita, Athanasios G.Tzioufas, Dimitrios I.Fotiadis,"

- Medical data quality assessment: On the development of an automated framework for medical data curation", *Computers in Biology and Medicine*, vol.107, pp.270-283, April 2019.
- [11] A.Idri, H.Benhar, J.L.Fernández-Alemán, I.Kadi, "A systematic map of medical data preprocessing in knowledge discovery", *Computer Methods and Programs in Biomedicine*, vol.162, pp.69-85, August 2018.
 - [12] Danilo Schmidt, Klemens Budde, Daniel Sonntag, Hans-Jürgen Profitlich, Matthias Ihle, Oliver Staack, "A novel tool for the identification of correlations in medical data by faceted search", *Computers in Biology and Medicine*, vol.85, pp.98-105, 1 June 2017.
 - [13] Hong Zhang, Wandong Ni, Jing Li, Youlin Jiang, Kunjing Liu, Zhaohui Ma, "On standardization of basic datasets of electronic medical records in traditional Chinese medicine", *Computer Methods and Programs in Biomedicine*, vol.174, pp.65-70, June 2019.
 - [14] Jia Su, Jinpeng Hu, Jingchi Jiang, Jing Xie, Yang Yang, Bin He, Jinfeng Yang, Yi Guan, "Extraction of risk factors for cardiovascular diseases from Chinese electronic medical records", *Computer Methods and Programs in Biomedicine*, vol.172, pp.1-10, April 2019.
 - [15] Nivethitha Somu, M. R. Gauthama Raman, Kannan Kirthivasan, V. S. Shankar Sriram, "Hypergraph Based Feature Selection Technique for Medical Diagnosis", *Journal of Medical Systems*, 40:239, November 2016.
 - [16] S. E. Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, and J. J. Struijk, "Acoustic Features for the Identification of Coronary Artery Disease," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2611-2619, Nov. 2015.
 - [17] J. Xu and K. Suzuki, "Max-AUC Feature Selection in Computer-Aided Detection of Polyps in CT Colonography," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 585-593, March 2014.
 - [18] J. Tang and S. Zhou, "A New Approach for Feature Selection from Microarray Data Based on Mutual Information," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 13, no. 6, pp. 1004-1015, 1 November 2016.
 - [19] S. Nijima and Y. Okuno, "Laplacian Linear Discriminant Analysis Approach to Unsupervised Feature Selection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 6, no. 4, pp. 605-614, Oct.-Dec. 2009.
 - [20] T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams, and I. C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," in *Proceedings of the IEEE*, vol. 106, no. 4, pp. 690-707, April 2018.
 - [21] T. Sekiya, A. Watanabe and M. Saito, "The use of modified constellation graph method for computer-aided classification of congenital heart diseases," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 8, pp. 814-820, Aug. 1991.
 - [22] Md. Zahangir Alam, M. Saifur Rahman, M. Sohel Rahman, "A Random Forest based predictor for medical data classification using feature ranking", *Informatics in Medicine Unlocked*, vol.15, 100180, 2019
 - [23] Ali Kalantari, Amirrudin Kamsin, Shahaboddin Shamshirband, Abdullah Gani, Hamid Alinejad-Rokny, Anthony T.Chronopoulos, "Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions", *Neurocomputing*, vol.276, pp.2-22, 7 February 2018.
 - [24] Richard Van Praagh, "Diagnosis of complex congenital heart disease: Morphologic-anatomic method and terminology", *CardioVascular and Interventional Radiology*, vol.7, no.3-4, pp 115-120, June 1984.
 - [25] Julio A. Barcena, James C. Fang, "Diagnosis and treatment of sleep apnea in heart disease", *Current Treatment Options in Cardiovascular Medicine*, vol.9, no.6, pp 501-509, December 2007.
 - [26] Rajakumar, Boothalingam, "Optimization using lion algorithm: a biological inspiration from lion's social behavior", *Evolutionary Intelligence*, vol.11, no. 1-2, pp.31-52, 2018.
 - [27] A.H. Gandomi, X.-S. Yang, S. Talatahari, and A.H. Alavi, "Firefly algorithm with chaos", *Commun Nonlinear Sci Numer Simulat*, vol. 18, pp. 89-98, 2013.
 - [28] John Mc Call, "Genetic algorithms for modelling and optimisation", *Journal of Computational and Applied Mathematics*, vol. 184, no. 1, pp. 205-222, 2005.
 - [29] M.E.H.Pedersen and A.J.Chipperfield, "Simplifying Particle Swarm Optimization", *Applied Soft Computing*, vol. 10, no. 2, pp. 618-628, 2010.
 - [30] Seyedali Mirjalili and Andrew Lewis, "The Whale Optimization Algorithm", *Advances in Engineering Software*, vol. 95, pp.51-67, 2016