






# A Model for Heart Disease Prediction Using Feature Selection with Deep Learning

Vaishali Baviskar<sup>1</sup> , Madhushi Verma<sup>2</sup> , and Pradeep Chatterjee<sup>3</sup> 

<sup>1</sup> GHRIET, Pune, India

[vaishali.baviskar@raisoni.net](mailto:vaishali.baviskar@raisoni.net)

<sup>2</sup> Bennett University, Greater Noida, India

[madhushi.verma@bennett.edu.in](mailto:madhushi.verma@bennett.edu.in)

<sup>3</sup> Tata Motors, Pune, India

[pchats2000@yahoo.com](mailto:pchats2000@yahoo.com)

**Abstract.** The heart disease is considered as the most widespread disease. It is challenging for most of the physicians to diagnose at an early stage to avoid the risk of death rate. The main objective of this study involves the prediction of heart disease by using efficient techniques based on feature selection and classification. For feature selection, the enhanced genetic algorithm (GA) and particle swarm optimization (PSO) have been implemented. For classification, the recurrent neural network (RNN) and long short term memory (LSTM) has been implemented in this study. The data set used is the Cleveland heart disease data set available on UCI machine learning repository, and the performance of the proposed techniques has been evaluated by using various metrics like accuracy, precision, recall and f-measure. Finally, the results thus obtained have been compared with the existing models in terms of accuracy. It has been observed that LSTM when combined with PSO showed an accuracy of 93.5% whereas the best known existing model had an accuracy of 93.33%. Therefore, the proposed approach can be applied in the medical field for accurate heart disease prediction.

**Keywords:** Heart disease prediction · Deep neural network · Long short term memory · Recurrent neural network · Genetic algorithm · Particle swarm optimization

## 1 Introduction

In the research field, heart disease has created a lot of serious concerns, and the significant challenge is accurate detection or prediction at an early stage to minimize the risk of death. According to World Health Organization (WHO) [1], the medical professionals have predicted only 67% of heart diseases correctly and hence there exists a vast research scope in the area of heart disease prediction. A lot of technicalities and parameters are involved in predicting the diseases

accurately. Various machine learning, deep learning algorithms and several optimization techniques have been used to predict the heart-disease risk. All these techniques mainly focus on the higher accuracy which shows the importance of correct prediction of heart disease. It would be helpful for the doctors to predict the heart disease at an early stage and save millions of life from death [2]. For temporal sequenced data, the recurrent neural network (RNN) models are best suited and for sequenced features, several variants have been chosen. In various sequence based tasks like language modelling, handwriting recognition, and for other such as tasks, long short term memory (LSTM) has been used, which shows an impressive performance [3,4]. For better performance, evolutionary algorithms (EAs) are used for model optimization. The evolutionary algorithm related to self-adaptability based on population is very useful in case of feature selection and extraction. The EAs used in the recent year include ant colony optimization (ACO), particle swarm optimization (PSO) and genetic algorithm (GA). The GA is considered as a stochastic method for optimization and global search, which is very helpful in handling the medical data. The possible solutions are obtained from set of individuals using GA. GA which are generally used to create solutions with a better quality for global search and optimization are based on the mutation, crossover and selection operators. The PSO-a meta heuristic algorithm is considered in this study due to its simplicity and ease implementation. It uses only few parameters and required few numbers of parameters tuning. The PSO exhibits the information sharing mechanism and population-based methods, and hence it extended from single to multi-objective optimization. It has been successfully applied in the medical field for heart disease prediction and recorded good performances [5,6]. The main contribution of this study involves,

- Improve the accuracy of the prediction of heart disease in human using efficient feature selection and classification methods.
- Implementing the GA and PSO for efficient feature selection.
- Implementing the RNN and LSTM to improve an accuracy for heart disease prediction.
- Compared performance of the proposed method with the existing techniques in terms of an accuracy, precision, recall and f-measure.

The remaining organization of the paper is as follows: Sect. 2 includes the literature survey of the existing research work related to feature selection techniques and deep learning classification methods for heart disease prediction. Section 3 discusses the implementation process of the GA and PSO optimization algorithm and LSTM and RNN classification. Section 4 discusses the performance analysis of the proposed work. The conclusion has been presented in Sect. 5.

## 2 Related Work

In [7], researchers proved that optimization algorithms are necessary for an efficient heart disease diagnosis and also for their level estimation. They used sup-

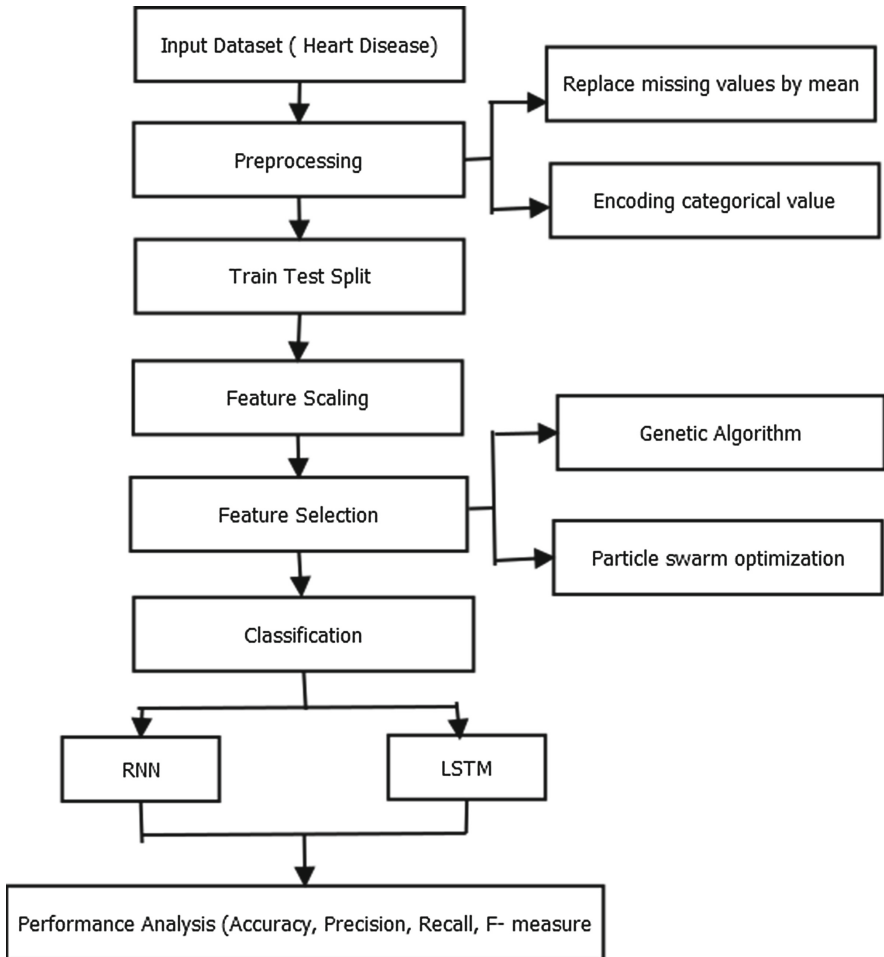
port vector machine (SVM) and generated an optimization function using the GA for the selection of more substantial features to identify the heart disease. The data set used in this research is a Cleveland heart disease database. G. T. Reddy et al. developed an adaptive GA with fuzzy logic design (AGAFL) in [8], which in turn helps the medical practitioners for heart disease diagnose at an early stage. Using the hybrid AGAFL classifier, the heart disease has been predicted, and this research has been performed on UCI heart disease data sets. For diagnosing the coronary artery disease, usually angiography method is used, but it shows significant side effects and highly expensive. The alternative modalities have been found by using the data mining and machine learning techniques stated in [9], where the coronary artery disease diagnosis is done with the more accurate hybrid techniques with increased performance of neural network and used GA to enhance its accuracy. For this research work, Z-Alizadeh Sani data set is used and yields above 90% values in specificity, accuracy and sensitivity. In [10], researchers proposed recurrent trained recurrent fuzzy neural network (RFNN) based on GA for heart disease prediction. The data set named UCI Cleveland heart disease is used. From the testing set, 97.78% accuracy has been resulted. For large data related to health diagnosis, the machine learning has been considered as an effective support system. Generally to analyze this kind of massive data more execution time and resources have required. Effective feature selection algorithm has been proposed by J. Vijayashree et al. in [11] to identify the significant features which contribute more in disease diagnosis. Hence to identify the best solution in reduced time the PSO has been implemented. The PSO also removes the redundant and irrelevant features in addition to selecting the important features in the given data set. Novel fitness function for PSO has been designed in this work using the support vector machine (SVM) to solve the optimal weight selection issue for velocity and position of particle's updation. Finally, the optimization algorithms show the merit of handling the difficult non-linear problems with adaptability and flexibility. To improve the heart disease classification quality, the Fast correlation based feature selection namely (FCBF) method used in [12] by Y. Khourdifi et al. to enhance the classification of heart disease and also filter the redundant feature. The classification based on SVM, random forest, MLP, K-Nearest neighbor, the artificial neural network optimized using the PSO mixed with an ant colony optimization (ACO) techniques, have been applied on heart disease data set. It resulted in robustness and efficacy by processing the heart disease classification. By using data mining and artificial intelligence, the heart disease has been predicted but for lesser time and cost in [13], which focused on PSO and neural network feed forward back propagation method by using the feature ranking on the disease's effective factors presented in Cleveland clinical database. After evaluating the selected features, the result shows that the proposed classified methods resulted in best accuracy. In [14], for the risk prediction of diseases, machine learning algorithm plays a major role. The prediction accuracy influenced by attribute selection in the data set. The performance metric of Mathew's correlation co-efficient has been considered. For attribute selection performance, the altered PSO has been

applied. N. S. R. Pillai et al. in [15] using the deep RNNs the language model like technique demonstrated to predict high-risk diagnosis patients (prognosis prediction) named as PP-RNNs. Several RNNs used by this proposed PP-RNN for learning from the patient's diagnosis code to predict the high risk disease existences and achieved a higher accuracy. In [16], M. S. Islam et al. suggested grey wolf optimization algorithm (GWO) combined with RNN, which has been used for predicting medical disease. The irrelevant and redundant attributes removed by feature selection using GWO. The feature dimensionality problem avoided by RNN classifier in which different diseases have been predicted. In this study, UCI data sets used and enhanced an accuracy in disease prediction obtained from Cleveland data set. From the structured and unstructured medical data, deep learning techniques exhibited the hidden data. In [17], researchers used the LSTM for predicting the cardio vascular disease (CVD) risk factors, and it generally yields better Mathew's correlation co-efficient (MCC) as 0.90 and accuracy as 95% compared with the existing methods. Compared with other statistical machine learning algorithms, the LSTM based proposed module shows best performance in the CVD risk factors' prediction. Based on novel LSTM deep learning method in [18], helped in predicting the heart failure at an early stage. Compared with general methods like SVM, logistic regression, MLP and KNN, the proposed LSTM method shows superior performance. Due to mental anxiety also CVD occurs, which may increase in COVID-19 lock down period. In [19], researchers proposed an automated tool which has used RNN for health care assistance system. From previous health records of patients for detecting the cardiac problems, the stacked bi-directional LSTM layer has been used. Cardiac troubles predicted with 93.22% accuracy from the obtained experimental results. In [21], Senthilkumar Mohan et al. proposed a hybrid machine learning technique for an effective prediction of heart disease. A new method which finds major features to improve the accuracy in the cardiovascular prediction with different feature's combinations and several known classification techniques. Machine learning techniques were used in this work to process raw data and provided a new and novel discernment towards heart disease. The challenges are seen in existing studies exhibited as,

- In the medical field, the challenging requirement is, training data in a large amount is necessary to avoid the over-fitting issue. Towards the majority samples, predictions are biased if the data set is imbalanced and hence over-fitting occurs.
- Through the tuning of hyper parameters such as activation functions, learning rates and network architecture, the deep learning algorithms are optimized. However, the hyper-parameters selection is a long process as several values are interdependent, and multiple trials are required.
- Significant memory and computational resources are required for timely completion assurance. Also, need to improve an accuracy of Cleveland heart disease data set using deep learning with feature selection techniques.

### 3 Methodology

The main purpose of this study is to predict the heart disease in human. The proposed workflow is shown in Fig. 1, which starts with the collection of dataset, data pre-processing, implementing the PSO and GA significantly for feature selection and for classification, RNN and LSTM classifiers used. At last, the proposed model is evaluated with respect to accuracy, precision, recall and f-measure. This section describes the workflow of the proposed study.



**Fig. 1.** Heart rate prediction proposed flow with RNN and LSTM classification

### 3.1 Preprocessing the Data

The data set contains 14 attributes and 303 records, where 8 are categorical and 6 are numeric attributes. Attribute of a class is zero for normal and one for having heart disease. Also, 6 records had been missing values. For pre-processing the two strategies is followed. By using features mean value; the missing values are replaced. Further the string values are changed into numerical categorical values. After filtering out such records, 297 were complete records available for the heart disease prediction.

### 3.2 Enhanced GA and PSO for Feature Selection

#### Genetic Algorithm (GA)

The GA is considered as the stochastic search algorithm that imitates natural evolutionary process using the operators which imposed upon the population. The GA algorithm used two major operators such as crossover operator and mutation operator. For individual's mating in parent population, the crossover operator is used whereas the characteristics of individual's randomly changed and diverse offspring resulted for the mutation operator. In the following algorithm 1, the offspring made systematic replacement of the generated parents. In nature the crossover of single point symmetric and through bit flipping the mutation is achieved. The expression of a minimization problem is,

$$fit = \alpha E(C) + \beta \frac{|sf|}{|Af|} \quad (1)$$

where,  $E(C)$  is the classifier's error rate,  $sf$  is the selected feature subset length and available features total count is the  $Af$ , the parameters used to control feature reduction and classification accuracy weights  $\beta$  is  $1 - \alpha$  and  $\alpha \in [0,1]$ .

#### Selection

It selected a portion of population for next-generation breed. Based on the measured fitness values using Eq. (1) the selection is generated.

#### Crossover

For further breeding, randomly selected two parents from the previously selected pool. Until the suitable population size reached, the process is continued. At only one point, the crossover taken place and this is the parent solution's mid-point. The crossover probability parameter is  $prob_c$  which controls the crossover frequency.

#### Mutation

Selected the random solutions from the chosen candidates for breeding and on these, the bit flipping has been carried out. A diverse group of solutions arise, which keeps various characteristics of their parents. The mutation probability parameter is  $Prob_m$  which controls the mutation's frequency.

**Table 1.** Algorithm 1

1. Set the initial population-size value $N$ , $n_{site}$ , $Prob_c$ , $Prob_m$ , $max_{it}$
2. Initializing the population at random by $x_i = (x_{i1}, x_{i2} \dots x_{iD})$
3. Assess fitness of every solution using Eq. (1)
4. $prevfit \leftarrow fit$
5. set counter=0
6. calculate fitness $fitsort = sort(fit(x))$
7. $x_{val} = \frac{N_{Top}}{2}(fitsort)$
8. for $j$ in $range(0, N/2)$ do
9. $ij = floor(\frac{N}{2} * random) + 1$
10. $ik = floor(\frac{N}{2} * random) + 1$
11. if $(Prob_c > random)$ then
12. $[x_{newval}(ij), x_{newval}(ik)] = crossover(x_{val}(ij), x_{val}(ik))$
13. Assess fitness functions for each of the new solutions using $f(x_i)$ Eq. (1)
14. if $(fit < prevfit)$ then
15. $x = x_{newval}$
16. $prevfit = fit$
17. end if
18. end if
19. if $(Prob_m > random)$ then
20. select $n_{site}$ no of random sites to mutate using
21. $il = floor(\frac{N}{2} * random) + 1$
22. $x_{newval}(il) = mutate(x_{val}(il))$
23. Assess fitness of every solution
24. if $(fit < prevfit)$ then
25. $x = x_{newval}$
26. $prevfit = fit$
27. end if
28. end if
29. $x =$ Combine $x_{val}$ and $x_{newval}$
30. end for
31. $[fit_{best}, i] = min(fit)$
32. $Sol_{best} = x(i)$
33. counter= counter+1
34. until $(counter < max_{it})$
35. Return $Sol_{best}$

GA algorithm 1 from Table 1 is described as, at first initialize the population size  $N$ ,  $n_{site}$ ,  $Prob_c$ ,  $max_{it}$  values. Then for each solution, initialize the population randomly as  $x_i = (x_{i1}, x_{i2} \dots x_{iD})$ . The following calculations are repeated

until the ending criteria is seen, i) evaluate the fitness value using  $f(x_i)$  ii) breeding population selected as  $x_{val} = N_{\frac{top}{2}}(fitsort)$  iii) Taken random value and its higher than  $Prob_c$ , random sample mutation from  $x_{val}$  is taken iv) update the enhanced new solution with existing solution v) Taken random value and its higher than  $Prob_m$ , random sample mutation from  $x_{val}$  is taken vi) update the enhanced new solution with existing solution vii) combination of  $x_{val}$  and  $x_{newval}$  generated and it is considered a new solution and finally global best solution is produced considered as best found solution.

### Particle Swarm Optimization (PSO)

The PSO is a population-based search technique derived from the exchange of information through birds. At first in PSO, initialize the particles' random population and based on their other particles interchange, these particles are moved with certain velocity in the population. At each iteration out of all particles, the personal, best and global best achieved and all the particles of velocity updated based on this information. To the personal, best and global, the weights are given by certain variables. The following algorithm 2 used specified transfer function type k is used for alteration of endless value to binary value, which is a substitute to basic hyperbolic tangent function. By the dimensional vector D, every particle is represented, and with every individual value, which is being binary are initialized randomly,

$$x_i = (x_{i1}, x_{i2} \dots x_{iD}) \in A_s \quad (2)$$

Where  $A_s$  is the search space which is available by dimensional vector D, the velocity  $v_i$  is represented and initialized to 0.

$$v_i = (v_{i1}, v_{i2} \dots v_{iD}) \in A_s \quad (3)$$

By each particle retained, the best personal position recorded as,

$$p_i = (p_{i1}, p_{i2} \dots p_{iD}) \quad (4)$$

From Table 2, the PSO algorithm described as, at first the swarm size values N, acceleration constant  $A_{c1}, A_{c2}, w_{max}, w_{min}, v_{max}, maxit$  are initialized. As in Eq. (2) and Eq. (3), the population is randomly initialized and velocity vectors are initialized respectively. The following calculations are repeated until the ending criterion is seen, i) inertia weight value w is updated, ii) using  $f(x_i)$  the each solution's fitness value is updated, iii) assigned the personal-best solution  $p_{best}$  and  $g_{best}$  as global test solution, iv) the velocity of each particle is formulated with respect to each iteration c, v) using the transfer function k, the continuous values are mapped into binary values and generate the new solutions. Finally, the global best is produced as best found solution.

### LSTM and RNN for Classification

A classification technique to predict the heart disease using the RNN and LSTM model is developed. The LSTM model is proposed at first by Hochreiter et al. in 1997 considered as special RNN model [20]. The RNN is a catch up to the current



**Table 2.** Algorithm 2

1. Set the initial swarm-size value N, Acceleration constant $A_{c1}, A_{c2}$ ,
2. $w_{max}, w_{min}, v_{max}, max_{it}$
3. Initialize population at random value x using Eq. (2) for every solution and v is the velocity vectors as dimensional D as 0 vectors as per in Eq. (3)
4. counter $\leftarrow$ c
5. set c=0
6. $w = w_{max} - c(\frac{w_{max}-w_{min}}{max_{it}})$
7. The fitness function for every solution is evaluated using $f(x_i)$ Eq. (1)
8. and assign the value for $p_{best}$ and $g_{best}$
9. for i in range(1,N) do
10. $v_i^{c+1} = wv_i^{c+1} + A_{c1}rand_1(p_{best}(c) - x_i^c) + A_{c2}rand_2(g_{best} - x_i^c)$
11. end for
12. update velocity v of particles
13. for i in range(1,N) do
14. for j in range(1,D) do
15. If $v(I,j) > v_{max}$ then
16. $v(i, j) = v_{max}$
17. end if
18. If $(v(I, j) < -v_{max})$
19. $v(i, j) = -v_{max}$
20. end If
21. $k = v_{max} \frac{1}{1+e^{(-v(i,j))}}$
22. If $(rand < k)$ then
23. x (I,j)=1
24. Else
25. $x(I, j) = 0$
26. end If
27. end for
28. end for
29. c=c+1
30. Until $c < max_{it}$
31. return $g_{best}$

hidden layer state to previous n-level hidden layer state to obtain the long-term memory. Basis of RNN network, the LSTM layers are added to valve node, which overcomes the RNN long term memory evaluation problems. Generally, LSTM includes three gates to original RNN network such as an input gate, forget gate and an output gate. The LSTM design key vision is to integrate data-dependent controls and non-linear to RNN cell is trained and assures that the objective

**Table 3.** RNN and LSTM Specification

Simple RNN layer (100 units)
LSTM layer (100 units)
Activation Layer - Softmax
Optimizer - Adam
Epoch - 100
Batch Size - 4

function gradient does not vanish based on the state signal. The specification of RNN and LSTM shown in Table 3.

GA and PSO algorithms with LSTM deep learning model are shown in Fig. 2 and Fig. 3. Here, GA and PSO are used as feature selection algorithms and LSTM is used as classifier to classify the patients into normal and abnormal class. Selected features are given as an input to classifier. The details of features selected are given in Table 6.

## 4 Results and Discussion

### 4.1 Dataset Description

In this proposed study, the Cleveland heart disease data set described is available on the UCI machine learning repository. This data set contains 303 instances, in which six instances exhibit missing attributes and 297 instances exhibit no missing data. In its original form, the data set contains 76 raw features. From Table 4, the experiments followed only 13 features. These instances have no missing values used in the proposed experiments.

**Table 4.** Dataset features description [22]

Feature no.	Feature description	Feature abbreviations	Features code
1	Age (Years) Numeric	AGE	F1
2	Sex	SEX	F2
3	Chest Pain Type	CPT	F3
4	Resting Blood Pressure	RBP	F4
5	Serum Cholesterol	SCH	F5
6	Fasting Blood Sugar	FBS	F6
7	Resting Electrocardiographic Results	RES	F7
8	Maximum Heart Rate Achieved	MHR	F8
9	Exercise Induced Angina	EIA	F9
10	Old Peak	OPK	F10
11	Peak Exercise Slope	PES	F11
12	Number of major vessels colored	VCA	F12
13	Thallium Scan	THA	F13

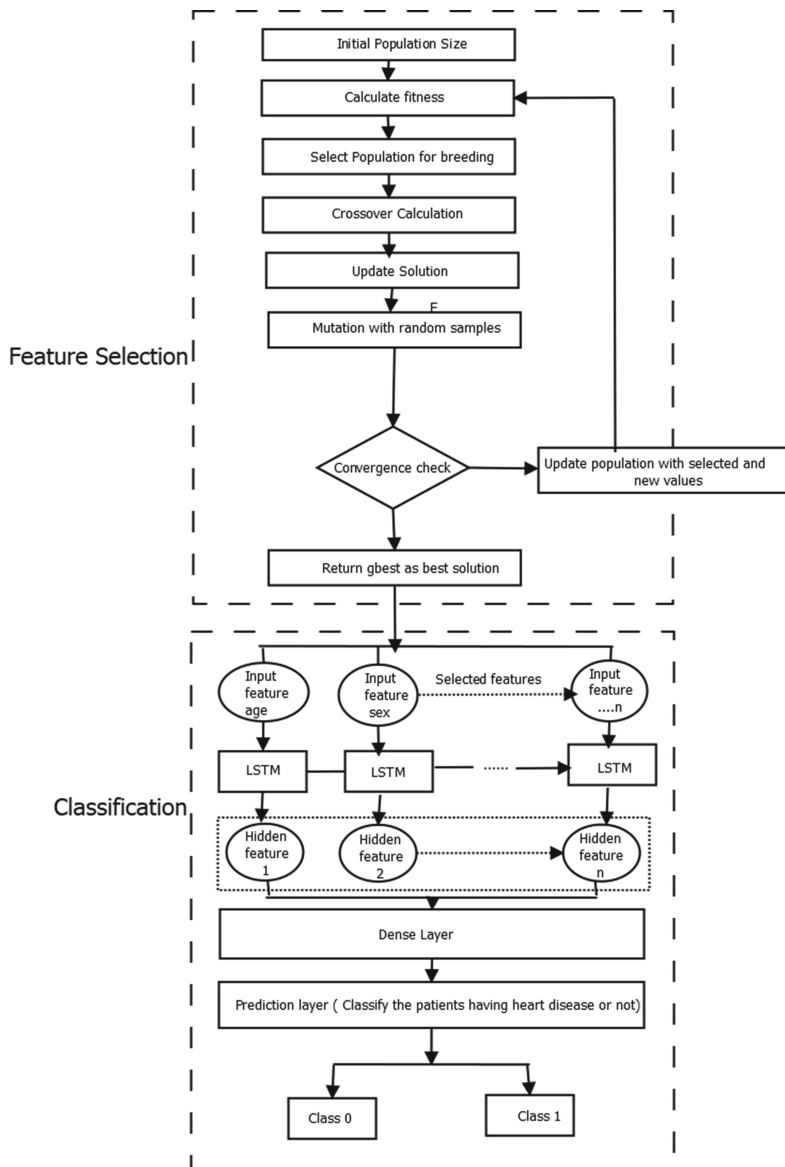


Fig. 2. GA with LSTM work flow

## 4.2 Performance Metrics

The proposed predictive model results are evaluated by performance metrics such as accuracy, precision, recall and f-measure. The formulations of all the metrics are,

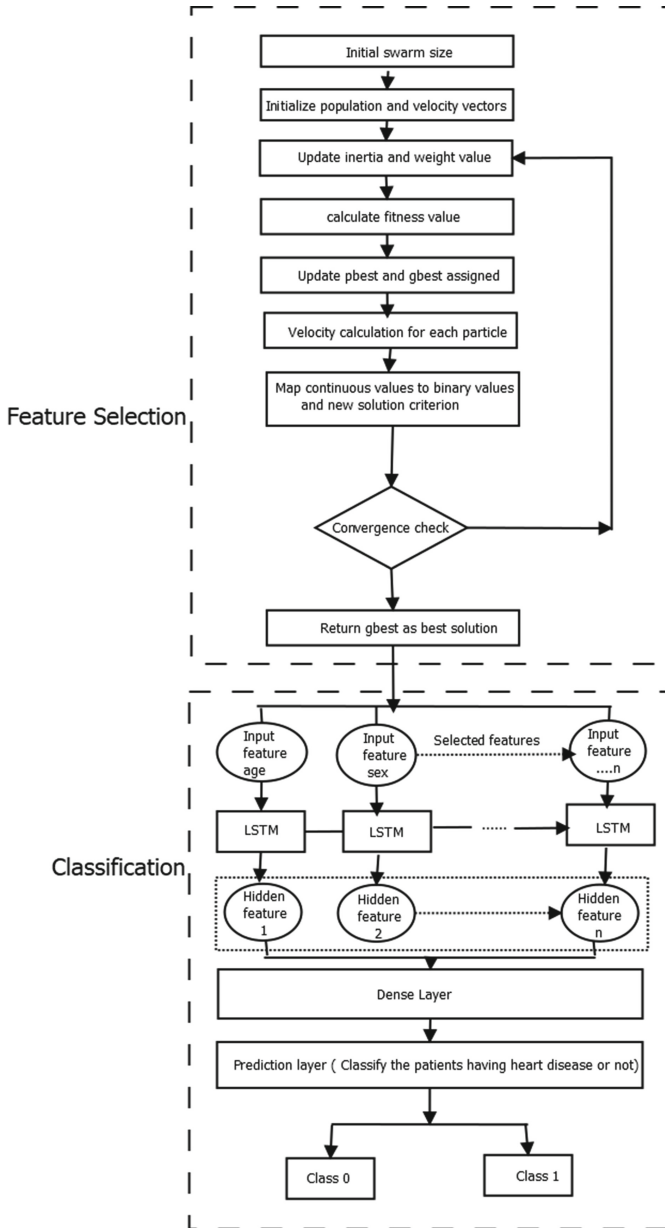


Fig. 3. PSO with LSTM work flow

**Accuracy:** The correctly classified in test data set shows in percentage values are termed as accuracy. The accuracy can be calculated based on the formula given in Eq. (5),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

**Precision:** While the correctly classified subjects showed by precision value. Precision is calculated by using the formula given in Eq. (6),

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

**Recall:** A recall is the proportion of related instances that have been recovered. Therefore, both accuracy and recall are based on an understanding of significance and measurement. It is estimated by the formula given in Eq. (7),

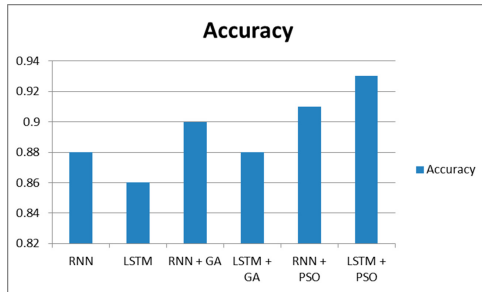
$$Recall = \frac{TP}{TP + FN} \tag{7}$$

**F-measure:** The method of F1 score is referred to as the harmonious mean of accuracy and recall. This can be computed with the aid of the formula given in Eq. (8),

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{8}$$

### 4.3 Comparison of Results

The following figures representing the performance metrics of the proposed method with respect to feature selection by GA and PSO and classification using RNN and LSTM shown below. Here, data set is split into 30% testing and 70% training. Out of 303, randomly 212 records have taken for training, and 61 records have taken for testing and predicted for normal (class 0 - negative class having no heart disease) and abnormal (class 1 - positive class having heart disease) class of heart disease.



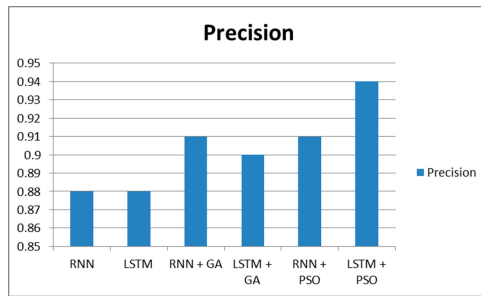
**Fig. 4.** Performance evaluation with an accuracy of model

From Fig. 4, it shows the results of the performance metric of accuracy of deep learning models, RNN and LSTM with and without feature selection algorithms

of GA and PSO. Here, all six models are compared and LSTM + PSO shows better accuracy of 93.5%. Out of 61 records tested, 57 predicted accurately where 25 records are from normal class, and 32 records are from abnormal class. Also, LSTM gives an accuracy in less time compared to RNN as shown in Table 5.

**Table 5.** Performance in time

Methods	Accuracy	Time
RNN	88.52	3.4672741889953613 s
LSTM	86.88	1.3537867069244385 s



**Fig. 5.** Performance evaluation with precision of model

From Fig. 5, it shows the results of the performance metric of precision of deep learning models, RNN and LSTM with and without feature selection algorithms of GA and PSO. Correctly classified of positive class patients’ percentage accuracy is shown. Here, all six models are compared and LSTM + PSO shows better performance of 94%. Out of 61 records tested, 34 records are predicted for abnormal class where 32 records are accurately predicted, and two records are from normal class but predicted wrongly as abnormal class.

From Fig. 6, it shows the results of the performance metric of recall of deep learning models, RNN and LSTM with and without feature selection algorithms of GA and PSO. Here, all six models are compared, and PSO shows better performance of 94% for RNN and LSTM classifier. Out of 61 records tested, 34 records are from abnormal class where 32 records are accurately predicted, and 2 are wrongly predicted as normal class.

From Fig. 7, it shows the results of the performance metric of f-measure of deep learning models, RNN and LSTM with and without feature selection algorithms of GA and PSO. Here, all six models are compared and LSTM + PSO shows better performance of 94%. It shows an average of precision and recall.

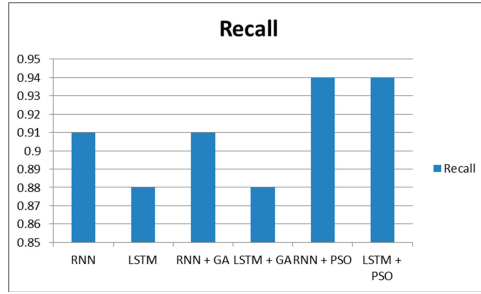


Fig. 6. Performance evaluation with recall of model

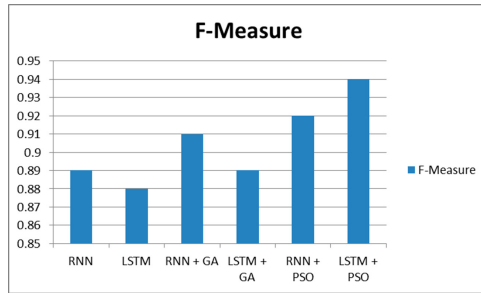


Fig. 7. Performance evaluation with F-measure of model

Table 6. Features accuracy

Methods	Selected feature	Feature count	Accuracy	Time
PSO	1110101011001	8	0.918033	92.51282
GA	1111111111010	11	0.901639	47.31271

From Table 6, the proposed method evaluation shows the PSO, and GA selected features. For PSO, the selected features’ count is 8 and shows an accuracy level as 91% and takes more time. While the GA selected features’ count is 11 and shows an accuracy level as 90% and takes lesser time compared with PSO. However, in terms of accuracy, the PSO shows better performance compared with GA.

From proposed Fig. 8, the evaluation performance for RNN is shown for GA and PSO features selected algorithms. It shows that, RNN with PSO shows the better performance compared to RNN with GA and without any feature selection. Also, accuracy is increased by 3% using PSO algorithm.

From proposed Fig. 9, the evaluation performance for LSTM is shown for GA and PSO features selected algorithms. It shows that, LSTM with PSO shows the better performance compared to LSTM with GA and without any feature selection. Also, accuracy is increased by 7% using PSO algorithm.

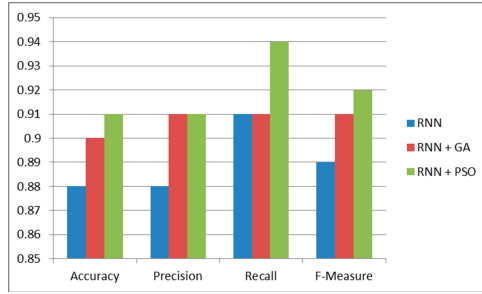


Fig. 8. Proposed method performance of RNN evaluation (With GA and PSO)

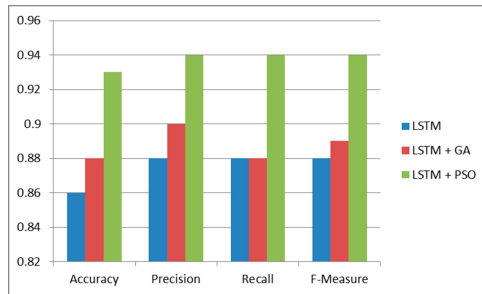


Fig. 9. Proposed method performance of LSTM evaluation (With GA and PSO)

Table 7. Comparison of Existing methods with Proposed method

Methods	Accuracy
DNN + $\chi^2$ Statistical model [22]	(K-fold)91.57
DNN + $\chi^2$ Statistical model [22]	(holdout) 93.33
RNN+GA ( Proposed method)	90
RNN+ PSO ( Proposed method)	92
LSTM+GA ( Proposed method)	90
LSTM+ PSO ( Proposed method)	93.5

From Table 7, it shows that by compared with the existing method the proposed method with LSTM + PSO shows higher accuracy for predicting the heart disease.

## 5 Conclusion

In this study, the efficient diagnosis approach has been developed for accurate prediction of heart disease. The proposed approach used enhanced GA and PSO for optimized feature selection from the heart disease data set. Further, the



classification has been achieved by using deep learning models such as RNN and LSTM. The proposed model has been evaluated using the accuracy, precision, recall and f-measure performance metrics. The obtained results show that the proposed method which implements LSTM with PSO yields an accuracy of 93.5% and slightly higher computational time due to the feature selection phase but leads to an accurate prediction of heart disease as compared to the existing methods. For other performance metrics like precision, recall and f-measure also LSTM + PSO shows better performance. In the future, it may be considered for enhancing the performance of the proposed model.

## References

1. Kirubha, V., Priya, S.M.: Survey on data mining algorithms in disease prediction. *Int. J. Comput. Trends Tech.* **38**, 124–128 (2016)
2. Sharma, H., Rizvi, M.: Prediction of heart disease using machine learning algorithms: a survey. *Int. J. Recent Innov. Trends Comput. Commun.* **5**, 99–104 (2017)
3. Choi, E., Schuetz, A., Stewart, W.F., Sun, J.: Using recurrent neural network models for early detection of heart failure onset. *J. Am. Med. Inform. Assoc.* **24**, 361–370 (2017)
4. Jin, B., Che, C., Liu, Z., Zhang, S., Yin, X., Wei, X.: Predicting the risk of heart failure with EHR sequential data modelling. *IEEE Access* **6**, 9256–9261 (2018)
5. Salem, T.: Study and analysis of prediction model for heart disease: an optimization approach using genetic algorithm. *Int. J. Pure Appl. Math.* **119**, 5323–5336 (2018)
6. Bobaoglu, I., Findik, O., Ulker, E.: A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector. *Expert Syst. Appl.* **37**, 3177–3183 (2010)
7. Gokulnath, C.B., Shantharajah, S.: An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Comput.* **22**, 14777–14787 (2019). <https://doi.org/10.1007/s10586-018-2416-4>
8. Reddy, G.T., Reddy, M.P.K., Lakshmana, K., Rajput, D.S., Kaluri, R., Srivastava, G.: Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol. Intell.* **13**, 185–196 (2020). <https://doi.org/10.1007/s12065-019-00327-1>
9. Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., Yarifard, A.A.: Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Comput. Methods Programs Biomed.* **141**, 19–26 (2017)
10. Uyar, K., İlhan, A.: Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. *Proc. Comput. Sci.* **120**, 588–593 (2017)
11. Vijayashree, J., Sultana, H.P.: A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. *Program. Comput. Softw.* **44**, 388–397 (2018). <https://doi.org/10.1134/S0361768818060129>
12. Khouridifi, Y., Bahaj, M.: Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *Int. J. Intell. Eng. Syst.* **12**, 242–252 (2019)

13. Feshki, M.G., Shijani, O.S.: Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network. In: Artificial Intelligence and Robotics (IRANOPEN) 2016, pp. 48–53 (2016)
14. Narasimhan, B., Malathi, A.: Altered particle swarm optimization based attribute selection strategy with improved fuzzy Artificial Neural Network classifier for coronary artery heart disease risk prediction. *Int. J. Adv. Res. Ideas Innov. Technol.* **5**, 1196–1203 (2019)
15. Pillai, N.S.R., Bee, K.K., Kiruthika, J.: Prediction of heart disease using RNN algorithm (2019)
16. Babu, S.B., Suneetha, A., Babu, G.C., Kumar, Y.J.N., Karuna, G.: Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network. *Period. Eng. Nat. Sci.* **6**, 229–240 (2018)
17. Islam, M.S., Umran, H.M., Umran, S.M., Karim, M.: Intelligent healthcare platform: cardiovascular disease risk factors prediction using attention module based LSTM. In: 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 167–175 (2019)
18. Maragatham, G., Devi, S.: LSTM model for prediction of heart failure in big data. *J. Med. Syst.* **43**, 111 (2019). <https://doi.org/10.1007/s10916-019-1243-3>
19. Bandyopadhyay, S.K., Dutta, S.: Stacked bi-directional LSTM layer based model for prediction of possible heart disease during lockdown period of COVID-19: bi-directional LSTM. *J. Adv. Res. Med. Sci. Technol.* **7**, 10–14 (2020). ISSN 2394-6539
20. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D: Nonlinear Phenom.* **404**, 132306 (2020)
21. Mohan, S., Thirumalai, C., Srivastava, G.: Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access* **7**, 81542–81554 (2019). <https://doi.org/10.1109/ACCESS.2019.2923707>
22. Ali, L., Rahman, A., Khan, A., Zhou, M., Javeed, A., Khan, J.A.: An automated diagnostic system for heart disease prediction based on  $\chi^2$  statistical model and optimally configured deep neural network. *IEEE Access* **7**, 34938–34945 (2019)