



# Innovative feature selection and classification model for heart disease prediction

Senthil Murugan Nagarajan<sup>1</sup> · V. Muthukumaran<sup>2</sup> · R. Murugesan<sup>2</sup> · Rose Bindu Joseph<sup>3</sup> · Munirathanam Meram<sup>4</sup> · A. Prathik<sup>5</sup>

Received: 23 October 2020 / Accepted: 22 July 2021 / Published online: 9 August 2021  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

## Abstract

Heart disease is a complex disease that affects a large number of people worldwide. The timely and accurate detection of heart disease is critical in healthcare, particularly in the field of cardiology. In this article, we proposed a system for diagnosing heart disease that is both efficient and accurate, and it is based on machine-learning techniques. The diagnosis of heart disease is found to be a serious concern, so the diagnosis has to be done remotely and regularly to take the prior action. In the present world, finding the prevalence of heart disease has become a key research area for the researchers and many models have been proposed in the recent year. The optimization algorithm plays a vital role in heart disease diagnosis with high accuracy. Important goal of this work is to develop a hybrid GCSA which represents a genetic-based crow search algorithm for feature selection and classification using deep convolution neural networks. From the obtained results, the proposed model GCSA shows increase in the classification accuracy by obtaining more than 94% when compared to the other feature selection methods.

**Keywords** Deep convolutional neural network · Crow search algorithm (CSA) · Genetic algorithm · Optimization · Healthcare analysis

---

✉ Senthil Murugan Nagarajan  
senthilgtc16@gmail.com

V. Muthukumaran  
muthu.v2404@gmail.com

R. Murugesan  
contactmurugu@gmail.com

Rose Bindu Joseph  
rosebindu@gmail.com

Munirathanam Meram  
munirathnam1986@gmail.com

A. Prathik  
prathik0505@gmail.com

<sup>1</sup> School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

<sup>2</sup> Department of Mathematics, School of Advance Sciences, REVA University, Bangalore, India

<sup>3</sup> Department of Mathematics, Christ Academy Institute for Advance Studies, Bangalore, India

<sup>4</sup> Department of Mathematics, Rajiv Gandhi University of Knowledge Technologies, R. K. Valley, Nuzvidu, Andhra Pradesh, India

## 1 Introduction

Heart disease (HD) is a major public health concern that has affected millions of people worldwide. HD is characterized by common symptoms such as shortness of breath, physical weakness, and swollen feet. Researchers are attempting to develop an efficient technique for the detection of heart disease, as current heart disease diagnosis techniques are ineffective in early detection for a variety of reasons, including accuracy and execution time. When modern technology and medical experts are unavailable, diagnosing and treating heart disease is extremely difficult. If the classifiers present in the deep convolution neural networks (DCNN) are of the same type, it is homogenous, and if they are of different types, then it is called as heterogeneous [7,25]. Every single DCNN is trained with their feature space and in some cases the features may have noise that contain the same data and

<sup>5</sup> Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

unwanted data. In such cases, the time taken to train is higher and also the false-positive rate is also higher.

To overcome this problem, feature selection technique is used. The feature selection technique is mainly used in the classification ensemble. While using the feature selection for the DCNN, it provides better results with the optimized features [25]. For obtaining the optimized feature subset, the Swarm and other algorithms are used. To perform the optimizing, the features algorithms such as PSO, genetic algorithms, support vector machine and machine-learning algorithm are used in large numbers. Artificial crow colony algorithm was developed by [6] to overcome the feature selection problems. Since it proposed, the ABC is used in various domains.

The GCSA algorithm provides better results with accuracy when compared with the genetic algorithm and Ant colony optimization algorithms. To improve the efficiency of classification, many algorithms have been proposed for classification based on patterns, different kinds of algorithms for feature selection and various ensemble classifiers for providing better accuracy rate [11,13]. The main disadvantage is these algorithms do not provide better results for the different kinds of datasets. This main goal of this work is to provide optimized feature selection with classifier assemblies for better accuracy rate. In this work, the optimized feature subset is obtained using the GCSA. The Cuckoo Search Algorithm is used to select the optimized features and the efficiency is calculated using a DCNN made up of Support vector machine, Naïve Bayes, random forest and decision tree [12]. This paper contains six sections: the Sect. 1 contains an introduction, the next comes related works, Sect. 3 is working on artificial crow colonies, the Sect. 4 is proposed artificial crow colony algorithm with DCNN, Sect. 4 is about the performance evaluation and finally the conclusion part.

## 2 Related works

Feature selection is done to find the best feature in a subset that can also be called as search in feature space. The feature selection is differentiated into two main categories, they are (i) filter method and (ii) wrapper method. In the filter approach, the filtration of features is done before the classifying the features, so they do not depend on the classification algorithm. The weight value is calculated for all the features in the dataset and thus compared with the original dataset [7]. In case wrapper method produces set of feature subset by addition and deletion of feature to get the feature subset, the accuracy is calculated for the feature subset to find the efficiency of the subset. From the various results obtained by the researcher, the wrapper method provides better results when compared with the filter method.

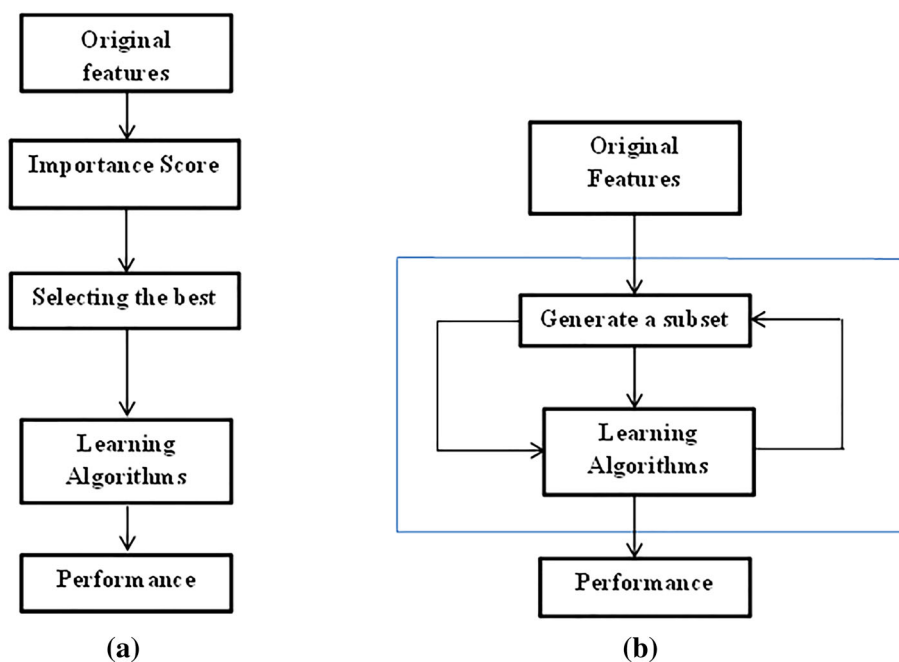
Many algorithms such as Info gain, filtered attribute, Ant colony optimization and bat algorithms are used for the process of feature selection. These are the evolutionary feature selections that are used by many researchers for the process of feature selection; the usage of the swarm optimization has increased in the last few years [20]. The model has been proposed with rough sets using Ant colony optimization for reducing the dimensionality in medical fields like dermatology. Ant colony algorithm with neural networks for the process of feature selection is also used. The particle swarm optimization algorithm can be used as both filter and wrapper method. A feature selection method is developed based on a wrapper model with bat algorithm with OPF classifier [18,28].

A new method states that ACO algorithm can also be used for feature selection, and that can also be used for image feature selection. ABC algorithm is an intelligent feature selection algorithm and used to solve the optimization problem with higher accuracy [15]. A model using Ant colony optimization is proposed and discussed the possibilities of using a meta-heuristic algorithm for selecting particular features that helps in obtaining higher accuracy. The concept for subspace clustering was proposed and it is based on a learning mechanism and it showed that each cluster in the dataset is not of the same dimension, so weight has been assigned to the each cluster in the dataset and it proves that feature selection is also required for clustering [3]. In the medical field, there is a huge volume of heterogeneous data such as medical record, test record, prescription, and scan reports, and that can be used in further stages of treatment [24]. The patterns discovered from various tests provide medical knowledge for the discoveries for example by finding the features, the disease can be identified. Scientists discovered that for using data mining algorithms in healthcare data, pre-processing of data has to be done [6,10].

The unwanted and repeated data can be removed from the dataset using the pre-processing techniques and this pre-processing is used to map the high-dimensional and low-dimensional data to reduce the time space required. The pre-processing can be done by two methods, (i) selecting the features and (ii) extracting the features, the feature extraction technique reduces the total number of features of the original dataset by combining the features of the dataset into a new subset. The feature selection selects the required features for performing the particular task. The feature selection works better with the medical data, because the originality of the feature remains unchanged so it is easy for a domain expert to select the required features. Wide range feature selection methods are available for finding the required features with accuracy [5,14].

A model called nine decision trees is developed, and it provides better results when compared to the decision tree and bagging algorithm. The main focus is the algorithm that are

**Fig. 1** Feature selection techniques: **a** filter method, **b** wrapper method



used for classification and the measuring the performance of DT algorithm and NB algorithm using the accuracy prediction technique and they proved that both algorithms provides better results after testing and in addition to that they stated that the decision tree is more cost effective than naïve Bayes when the dataset has less attributes and instance [17,22].

The authors in [19] presented a novel time series-based approach for the early prediction of increase in hypertension by analyzing the electrocardiograph holter signals. The authors in [21] proposed feature selection algorithm for selecting the suitable feature from the available dataset. The authors proposed genetic algorithm-based recursive feature elimination technique and shown a better outcome. The authors in [9] built a health monitoring system based on Internet of Things (IoT) and analyzed the Lamb waves to determine concrete structure of health. Coronato and Cuzocrea [2] proposed Dynamic Probabilistic Risk Assessment of Medical Information Systems for improving the medical device post market surveillance which is currently implemented as a wait for an incident activity.

The active feature selection model proposed, in this model, the instance of the features are actively selected. The author stated that each feature selection has its own advantages and disadvantages; the performance of the algorithm reduces while using a large dataset [1,8]. The authors proposed a new method using a decision tree with bagging and backward elimination strategy that is used to find the relationship between the chemometrics and its related paramedical industry. The model was proposed to discuss the requirement of feature selection in both supervised and unsupervised learning [23,26,27].

### 3 Feature selection

The important step in pre-processing is feature selection and it is used in different tasks of data mining algorithms for example pattern classification. If the subset of feature space is high, the feature selection selects only necessary features by removing the unwanted features in data space. Otherwise the unwanted data increases the time and complexity to compute, by adding the unwanted and repeated data in the process. In feature selection, important features are selected according to their importance, so the classification of features can be done easily without changing the original subset. Many researchers proved that classification done by features obtained from feature selection have a higher accuracy rate than classification done without feature selection (Fig. 1).

Many algorithms are developed for selecting the features. Feature selection is the same as pattern classification and they fall under two methods, filter technique and wrapper technique [4,16]. If the feature selection process is independent, then it is called a filter-based technique and it depends on the characters of the data. If the classifier is used, then it is called a wrapper approach; the features obtained from the wrapper method depend only on the classification algorithm used. Two different classifier algorithms provide two different feature subsets. The feature subset obtained from the wrapper method is more effective when compared to the filter method; the wrapper approach is a time-consuming process.

## 4 Proposed GCSA with DCNN

The GCSA is combined with the DCNN to provide a better solution to optimization problems. The DCNN is a combination of four algorithms namely DT, SVM, RF and NB algorithms. In the proposed model, GCSA algorithm, the artificial crow colony is to find the features and the subset of feature generator, to evaluate the feature subset obtained DCNN is used. The DCNN is used to find the feature that is suggested by the ABC model. The ABC helps the DCNN in constructing the subset of best features. The ABC algorithm and DCNN enhance the performance of both the algorithms (Fig. 2).

### 4.1 GCSA-based feature selector

The GCSA is an intelligent algorithm used for the feature optimization process and this colony and genetic algorithm increases the accuracy of the ensemble. The DCNN is combination of four algorithms such as support vector machine, naïve Bayes, decision tree and random forest, and with these algorithms, the ability of each feature in subset can be calculated,  $F_i$  in the feature subset. 10-fold technique is used to find the accuracy of available feature in the subset. Each employed crow is represented in binary string 0 or 1. The total number of features is the same as the length of the dataset, and it also represents feature selection done by the crow search. In the binary string, 1 represents the feature is selected and 0 represents feature is not selected. The total amount of onlooker and crow search is the same as the number of features in the dataset (Fig. 3).

After the completion of genetic process for the initial population, then the crow search is applied. The initialization of crows is the extraction of feature from the dataset. Initially,  $C_i$  is the feature based on crows that are arbitrarily located in a search space and given in following equation:

$$C_i = C - 1, C_2, \dots, C_n, \quad \text{where, } i = 1, 2, 3 \dots n. \quad (1)$$

Few initial solutions are used at the starting stage for the meta-heuristics optimization models for improving by monitoring the contrast solution simultaneously. The opposition point definition can be written as per the following equation:

$$\hat{c}_i = g_j + h_j - c_i. \quad (2)$$

The fitness function for the OCS can be calculated based on the basis function as per the following equation:

$$OC_i = \text{MAX}(\text{Accuracy}). \quad (3)$$

One of the flock crow is randomly chosen to form a novel position and the innovative position of the crow is obtained

### Algorithm 1 GCSA-based Feature Selection Algorithm

---

```

1: Feature set is taken as a initial population for the GA.
2: Evaluate fitness function for the optimal feature.
3: for each reproduction of optimal feature do
4:   Selecting chromosomes based on the fitness function of the feature.
5:   Crossover of the randomly generated feature set. Mutate child chromosomes with its mutation probability.
6:   New population generated based on GA
7:   Obtain opposition process for initial population
8:   Calculate fitness function for opposition processed population
9:   Calculate fitness function for Optimal solution from GA process
10:  Obtain best solution
11:  Calculate the objective function for best solution
12:  Generate new position
13:  if new population = feasible then
14:    Update Memory and Position
15:  else
16:    Stop with the current best population
17:  if Max iteration is obtained then
18:    return best solution
19:  else
20:    Repeat until the iteration process and go to step 12
21:  End
22: Best Feature set is selected as optimal solution

```

---

using the following equation:

$$PC^{i+iter+1} = \begin{cases} PC_{i+iter+1} + a_i \times cl^{i,iter} \\ \times (M^{j,iter} - PC^{i,iter}) & \text{if } a_j \geq APC^{j,iter} \\ \text{ran PC} & \text{otherwise.} \end{cases} \quad (4)$$

The current position and memory of the upgraded crow is processed based on the following equation:

$$M^{i,iter+1} = \begin{cases} PC^{i,iter} & c(PC^{j,iter+1}) > c(M^{j,iter}) \\ M^{i,iter} & \text{otherwise.} \end{cases} \quad (5)$$

The fitness value of the crow's new position is controlled to be the highest in the sustained position. Crow often updates storage space with new locations. If multiple iterations are implemented, the best memory location corresponding to the target will be addressed as the best filtering feature set solution. The suitability of the role, the health assessment is done through the function of each crow:

$$\text{fitness}_1(S) = \frac{\sum_{j=1}^m \text{accuracy}(S)}{m} \quad (6)$$

$$\text{fitness}_2(S) = \text{consensus}(S) \quad (7)$$

$$\text{fitness} = \frac{\text{fitness}_1(S) + \text{fitness}_2(S)}{2}. \quad (8)$$

The accuracy( $S$ ) is the predicted value in the ensemble classifier, consensus( $S$ ) represents the accuracy of classification on the  $S$  feature subset. The fitness can be evaluated

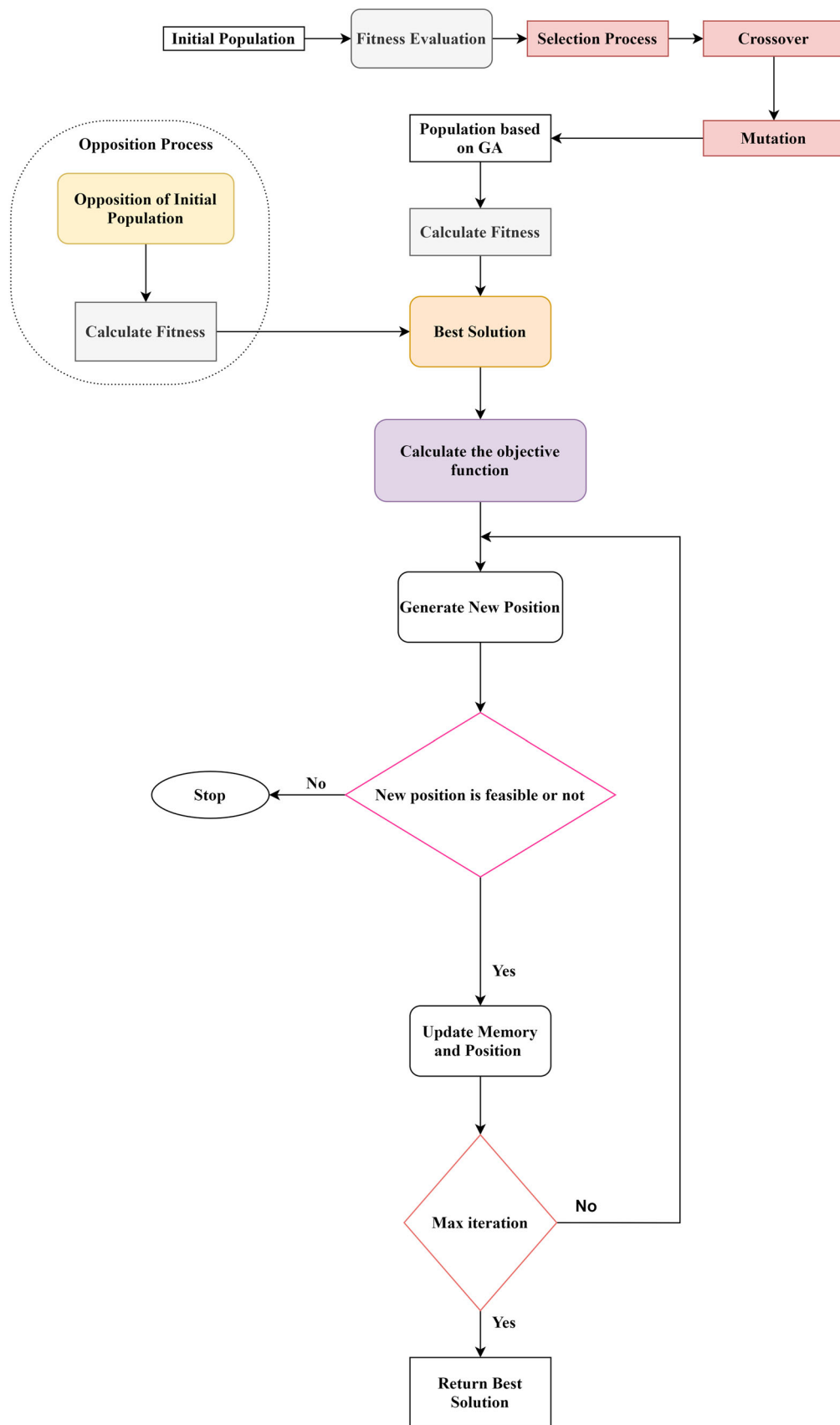
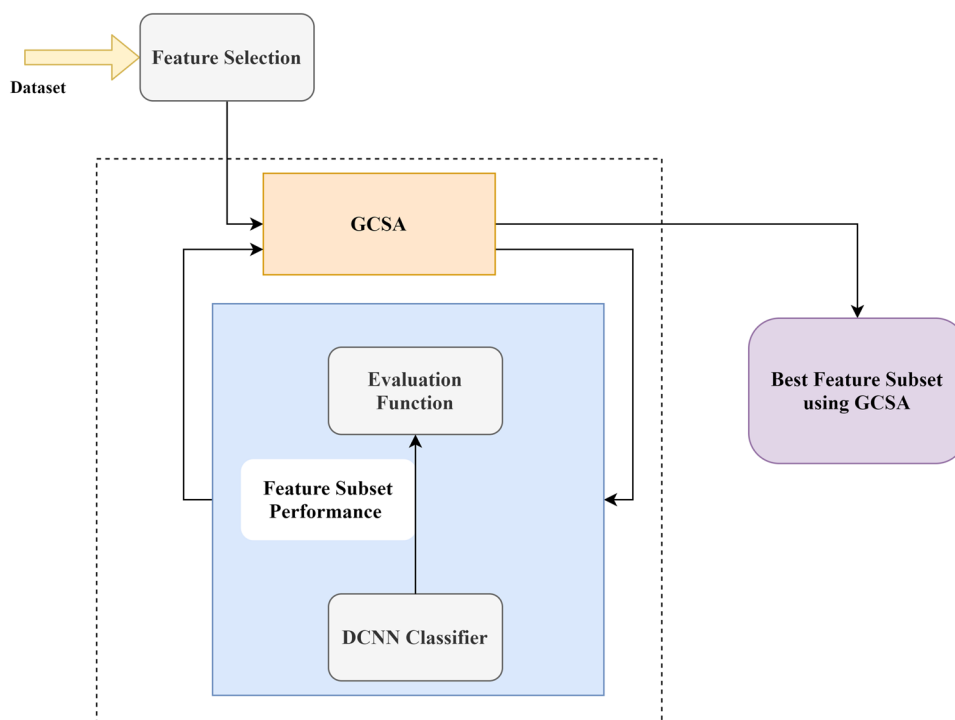


Fig. 2 Flowchart process of proposed algorithm

**Fig. 3** Proposed model for heart disease prediction



using mean accuracy and consensus, mean accuracy checks whether the features have the power for accurate classification, the DCNN tries for feature optimization. The mean accuracy helps in increasing the ability of feature subset generalization. The second phase of fitness evaluation is consensus; it finds whether the feature subset has optimality in classification for producing high consensus classification. The crow search passes the information to the onlooker crow and it checks the likelihood of feature selection using Eq. (7), the new solution given by onlooker crow is represented as  $V_i$ , using the mean and consensus value of the feature the crow search points out the feature selected by the onlooker crow. If the newly obtained value  $V_i$  is larger than  $X_i$ , then the crow search points to the feature in the feature subset that has crown previously selected and the new one. If the value of  $V_i$  is less compared to  $X_i$ , then crow search features will be used for further process, and the newly selected feature is omitted. The  $V_i$  is obtained using the following formula:

$$P_i = \frac{\text{fitness}}{\sum_{i=1}^m \text{fitness}} \quad (9)$$

$$V_i = X_i + \mu_i (X_i - X_j) \quad (10)$$

where  $X_i$  is the selected feature's accuracy,  $X_j$  is the accuracy of feature selected by onlooker crow.  $\mu_i$  is the randomly generated number in range (0, 1).

Therefore, when the crow search is allocated with a new feature, the onlooker crows make full use of it and a new

configuration for the subset is produced, after this process, all the features are used for forming a new feature subset, the available content in features tries to move to better feature subset configuration. If the improvement is not made in crow search then the employee becomes scout crow. Then, the new feature subset is assigned to the scout crow that is represented as follows:

$$X_{ij} = X_j^{\max} + \text{rand}(0, 1)(X_j^{\max} - X_j^{\min}) \quad (11)$$

where  $X_j^{\max}$  is the upper boundary value and  $X_j^{\min}$  is the lower boundary value.

The upper and lower boundaries and the same process have crown carried out till the stopping criteria are achieved to get the best features.

The obtained optimal solution is applied as input for the genetic algorithm and evaluated for the fitness of each feature and based on the chromosome, the child is selected as features and mutation process for that child-chromosome is selected and repeated until the given iteration value. From this, the GCSA can select the features based on their ranking, so the important feature is selected from the feature subset; the time consumption caused by the noisy and unwanted features is reduced. In case of large datasets, the classifier performance is reduced because of the huge number of features handled. In the GCSA algorithm, the features are selected based on their importance and classifier computation speed is enhanced.

**Table 1** Attributes of the dataset

S. no.	Features	Description
1.	Cp	Chest_pain types Value_1: typical angina Value_2: non-angina pain Value_3: atypical angina Value_4: asymptomatic
2.	Age	Age (years)
3.	Sex	Female (or) male
4.	Chol	Serum-cholesterol (md/dl)
5.	Trestbps	Blood pressure for resting (mmHg)
6.	Fbs	Fasting-blood pressure (120 mg/dl)
7.	Thalach	Heart rate attained (maximum)
8.	Oldpeak	Depression (ST)
9.	Restecg	Result of electrocardiographic
10.	Ca	No. of major vessels
11.	Slope	St-segment slope
12.	Thal	Status of heart
13.	Class	Attributes class label

**Table 2** Attributes of the dataset

S. no.	Dataset	Instance	Features	Class
1.	Heart	305	15	2
2.	Wisconsin cancer	697	11	2
3.	Lung cancer	31	55	2
4.	Diabetes	765	8	2
5.	Iris	148	5	3
6.	Pima Indian	765	9	2
7.	Lymphography_Dataset	146	18	4
8.	Hepatitis dataset	153	19	2
9.	Statlog heart disease	268	12	2
10.	Dermatology	364	31	6

**Table 3** Features selected and accuracy of GCSA

S. no.	Dataset	Total number of features	Features selected after using GCSAE	Accuracy of GCSA
1.	Heart-C dataset	15	8	89.10
2.	Dermatology dataset	32	23	97.54
3.	Lung cancer	55	26	81.30
4.	Pima Indian	9	5	89.40
5.	Hepatitis	18	9	80.01
6.	Iris	5	3	96.04
7.	Wisconsin breast cancer	10	5	97.97
8.	Lymphography	17	8	98.68
9.	Diabetes	8	6	83.16
10.	Statlog heart disease	12	6	88.30

## 4.2 DCNN

In this paper, the classifying algorithms such as DT, SVM, RF and NB are combined to form an ensemble classifier. In this work, crows search for the features, the features selected by each crow becomes the input to the classifier. The evaluation of features is done one by one and each feature has to be evaluated separately. The training of the subset is done using the four classifier algorithm and classifies the test subset. After classification is done, the cuckoo search algorithm calculates the mean accuracy and consensus of ensemble using formula. The average of mean accuracy and consensus is used to calculate the fitness of the features. For the process of selecting the best features, the fitness function is used.

## 5 Results and discussion

Dataset used implementation of proposed work and result obtained compared with other classifier algorithms are discussed in this phase.

### 5.1 Dataset

The performance of the proposed GCSA with DCNN is implemented and tested using ten different medical datasets. The ten different datasets are dermatology dataset, heart-C dataset, lung cancer dataset, pima Indian dataset, hepatitis disease dataset, Iris disease dataset, Wisconsin cancer dataset, Lymphography dataset, diabetes disease dataset and Statlog heart disease dataset and these kinds of datasets are used. The attributes of the dataset are given in figure. These datasets are used for DCNN and feature selection by many researchers so we have used these datasets for performance evaluation of proposed algorithms. This dataset is used because it has many attributes and instances so the accuracy of the proposed algorithm can be found easily (Table 1).

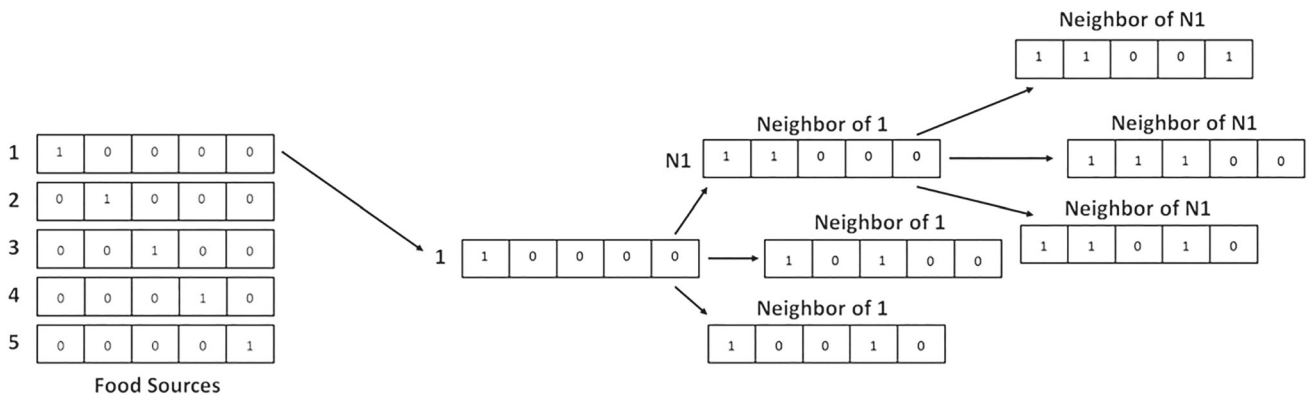
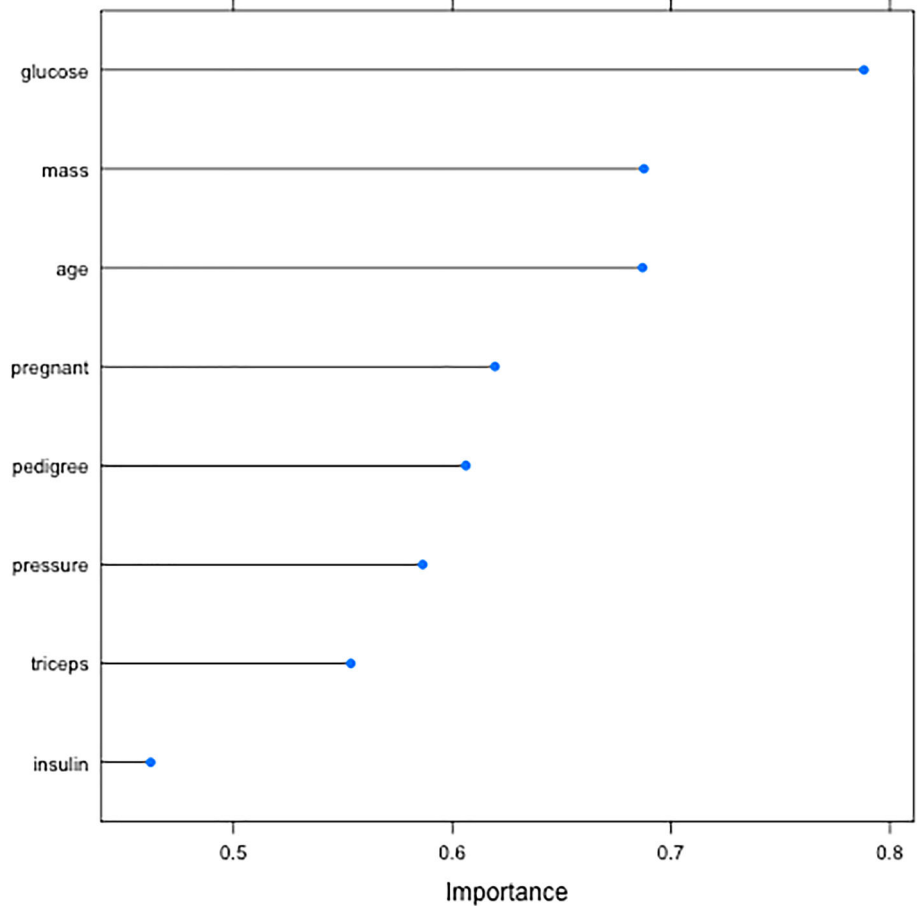


Fig. 4 Ranking features by importance [24]

Fig. 5 Feature selection using GCSA



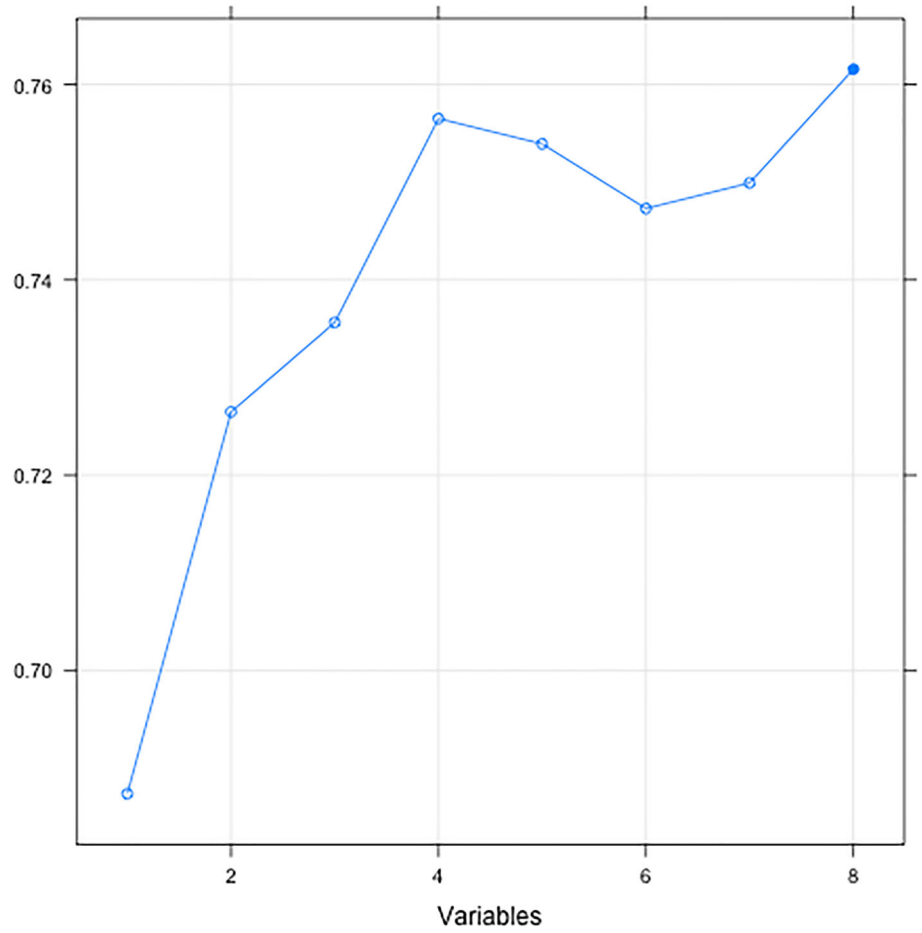
The performance of the proposed GCSA DCNN is implemented and tested using ten different medical datasets. The ten different datasets are dermatology dataset, heart-C dataset, lung cancer dataset, Pima Indian dataset, hepatitis dataset, Iris dataset, Wisconsin cancer dataset, lympho dataset, diabetes disease dataset and Statlog disease dataset being used (Table 2).

The parameters for selecting the best feature subset is set, the best feature subset is obtained after the predetermined

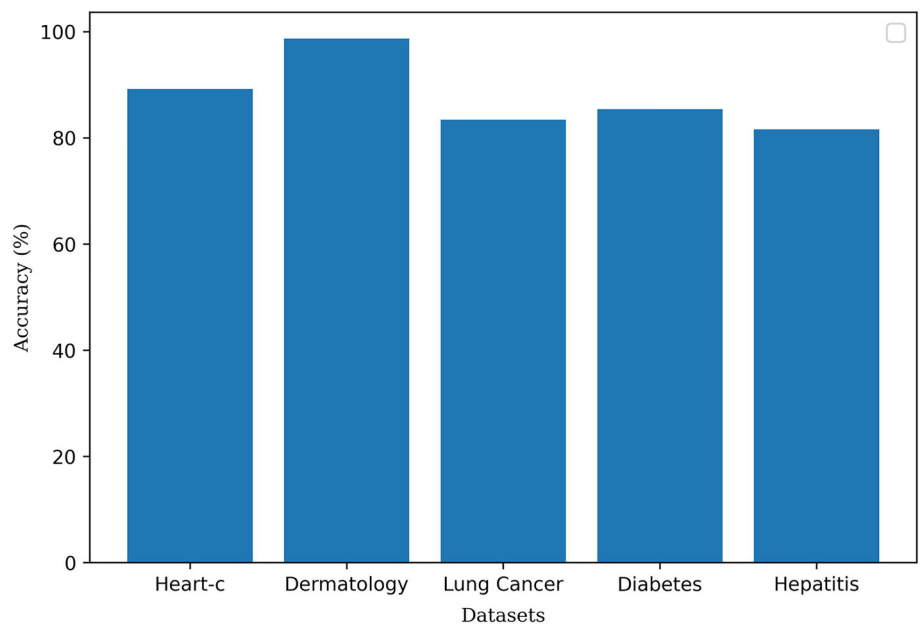
number of cycles. The employed crow passes the selected features to the DCNN after every iteration. The mean accuracy and consensus of ensemble classifiers are calculated using formula (4) and (5), the fitness of the feature subset is average of accuracy and consensus. The onlookers select the features with a probability based on fitness. The number of selected features and accuracy of classification is given in Table 3 (Figs. 4, 5, 6).



**Fig. 6** Accuracy of proposed GCSA DCNN [22]

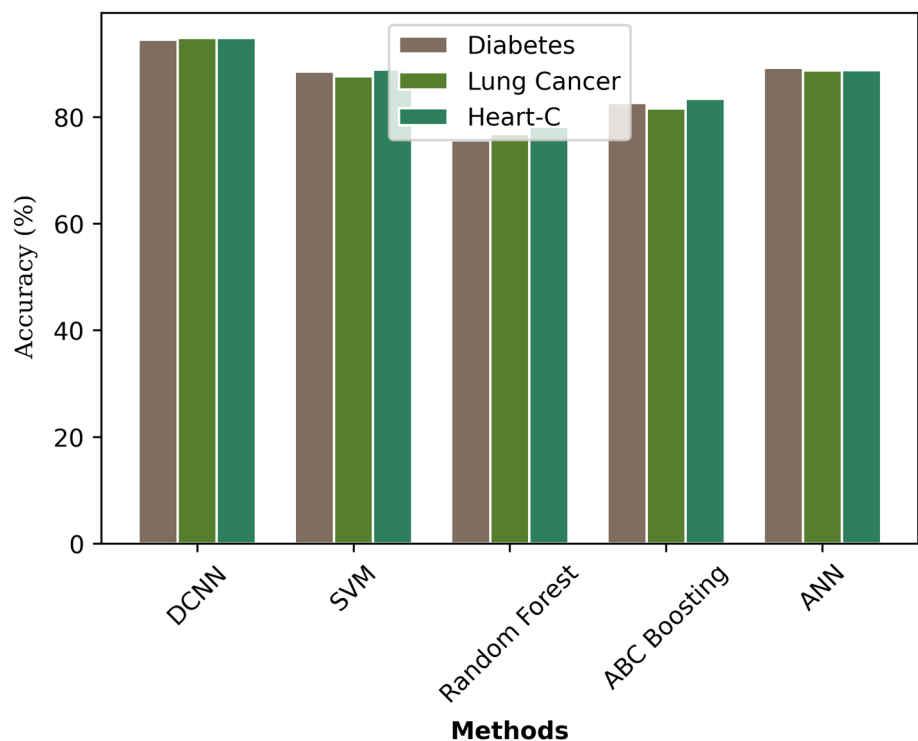


**Fig. 7** Comparison of various algorithms



**Table 4** Features selected and accuracy of GCSAE

S. no.	Datasets	Bagging with C4.5 method	Boosting with C4.5 method	ACO bagging	ACO boosting	GCSA with DCNN
1.	Heart-C dataset	77.87	77.56	87.56	87.56	91.78
2.	Dermatology disease dataset	96.20	96.50	97.55	97.55	97.12
3.	Lung cancer	84.20	85.45	77.34	77.34	88.23
4.	Pima Indian	78.34	75.10	89.45	87.45	89.34
5.	Hepatitis	74.50	74.28	79.30	77.80	80.09
6.	Iris	95.23	93.45	90.56	93.45	96.02
7.	Wisconsin breast cancer	95.78	96.47	78.45	78.12	96.10
8.	Lymphography	74.56	73.45	88.56	88.09	96.09
9.	Diabetes	79.90	84.12	83.45	84.12	83.50
10.	Statlog heart disease	80.30	81.45	80.35	82.07	83.45

**Fig. 8** Comparative analysis of different algorithms based on accuracy

Automatic feature selection models can be used to select the features. The caret package provides an automatic feature selection model called Recursive Feature Elimination. The RFE model is used in the Pima Indians diabetes dataset. Random forest is used for the iteration to evaluate the model. Figure 7 explains the graph plotted using ranking features by their importance, which gives a good resemblance of the feature selection.

The four DCNN is developed using the optimized feature subset obtained by the proposed GCSA algorithm. The classification accuracy of four classifiers is shown in Table 4. The performance of GCSA is compared with Ant colony optimization, C4.5 bagging and C4.5 boosting. A 10-fold

cross-validation algorithm is used to evaluate the accuracy of constructed classifiers. Classification accuracy rate of ten dataset increases rapidly after using the proposed artificial crow DCNN algorithm.

The performance of proposed GCSA with DCNN is compared with many other algorithms such as Decision tree, Support vector Machine, genetic algorithm, gain ratio, one attribute based, and filter attribute based, for the comparison, Statlog heart disease dataset is used. Graph is drawn to show good performance of the proposed GCSA–DCNN. The number of features selected by the proposed artificial crow colony algorithm is 8 features. The number of features selected by the different feature selection algorithm is given in the table.

The GCSA with DCNN algorithm performs better than all the bagging and boosting performed earlier (Fig. 8).

## 6 Conclusion

In this research, the proposed optimization function based on DCNN is introduced for enhancing the performance of ABC technique. ABC algorithm with DCNN-based feature selection algorithm identifies eight features. The resultant features are passed to the DCNN for finding the accuracy. The GCSA model obtained an accuracy of 88.78% for all original features and 95.34% for extracted features. The accuracy for the proposed algorithm is high when compared to other feature selection such as Decision tree, support vector machine, and artificial crow colony.

## References

- Ang JC, Mirzal A, Haron H, Hamed HNA (2015) Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM Trans Comput Biol Bioinform* 13(5):971–989
- Coronato A, Cuzzocrea A (2020) An innovative risk assessment methodology for medical information systems. *IEEE Trans Knowl Data Eng* :1–1. <https://doi.org/10.1109/tkde.2020.3023553>
- Ge Z, Song Z, Ding SX, Huang B (2017) Data mining and analytics in the process industry: the role of machine learning. *IEEE Access* 5:20590–20616
- Hira ZM, Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. *Adv Bioinform* 2015 :1–13. <https://doi.org/10.1155/2015/198363>
- Hu B, Dai Y, Su Y, Moore P, Zhang X, Mao C, Chen J, Xu L (2016) Feature selection for optimized high-dimensional biomedical data using an improved shuffled frog leaping algorithm. *IEEE/ACM Trans Comput Biol Bioinform* 15(6):1765–1773
- Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. *Appl Soft Comput* 8(1):687–697
- Karunyalakshmi M, Tajunisha N (2017) Classification of cancer datasets using artificial bee colony and deep feed forward neural networks. *Int J Adv Res Comput Commun Eng* 62:33–41
- Manogaran G, Alazab M, Saravanan V, Rawal BS, Shakeel PM, Sundarasekar R, Nagarajan SM, Kadry SN, Montenegro-Marin CE (2020) Machine learning assisted information management scheme in service concentrated IoT. *IEEE Trans Ind Inform* 17(4):2871–2879
- Misra D, Das G, Das D (2020) An IoT based building health monitoring system supported by cloud. *J Reliab Intell Environ* 6:141–152
- Muni Kumar N, Manjula R et al (2014) Role of big data analytics in rural health care—a step towards Svasth Bharath. *Int J Comput Sci Inf Technol* 5(6):7172–7178
- Murugan NS, Devi GU (2018) Detecting spams in social networks using ml algorithms—a review. *Int J Environ Waste Manag* 21(1):22–36
- Murugan NS, Devi GU (2018) Detecting streaming of twitter spam using hybrid method. *Wirel Pers Commun* 103(2):1353–1374
- Murugan NS, Devi GU (2019) Feature extraction using LR-PCA hybridization on twitter data and classification accuracy using machine learning algorithms. *Clust Comput* 22(6):13965–13974
- Nagarajan SM, Deverajan GG, Chatterjee P, Alnumay W, Ghosh U (2021) Effective task scheduling algorithm with deep learning for internet of health things (ioht) in sustainable smart cities. *Sustain Cities Soc* 71:102945
- Nagarajan SM, Muthukumar V, Murugesan R, Joseph RB, Muniathanam M (2021) Feature selection model for healthcare analysis and classification using classifier ensemble technique. *Int J Syst Assur Eng Manag*. <https://doi.org/10.1007/s13198-021-01126-7>
- Nagpal A, Gaur D (2015) ModifiedFAST: a new optimal feature subset selection algorithm. *J Inf Commun Converg Eng* 13(2):113–122
- Nalband S, Sundar A, Prince AA, Agarwal A (2016) Feature selection and classification methodology for the detection of knee-joint disorders. *Comput Methods Programs Biomed* 127:94–104
- Ng K, Ghoting A, Steinhubl SR, Stewart WF, Malin B, Sun J (2014) PARAMO: a parallel predictive modeling platform for healthcare analytic research using electronic health records. *J Biomed Inform* 48:160–170
- Paragliola G, Coronato A (2021) An hybrid ECG-based deep network for the early identification of high-risk to major cardiovascular events for hypertension patients. *J Biomed Inform* 113:103648
- Rani AS, Rajalaxmi RR (2015) Unsupervised feature selection using binary bat algorithm. In: 2015 2nd International conference on electronics and communication systems (ICECS). <https://doi.org/10.1109/ecs.2015.7124945>
- Rani P, Kumar R, Ahmed NM, Jain A (2021) A decision support system for heart disease prediction based upon machine learning. *J Reliab Intell Environ*. <https://doi.org/10.1007/s40860-021-00133-6>
- Saxena K, Sharma R et al (2015) Diabetes mellitus prediction system evaluation using c4. 5 rules and partial tree. In: 2015 4th International conference on reliability, infocom technologies and optimization (ICRITO) (trends and future directions). <https://doi.org/10.1109/icrito.2015.7359272>
- Shahana AH, Preeja V (2016) Survey on feature subset selection for high dimensional data. In: 2016 International conference on circuit, power and computing technologies (ICCPCT). <https://doi.org/10.1109/iccpct.2016.7530147>
- Shardlow M (2016) An analysis of feature selection techniques, vol 1. The University of Manchester, Manchester, pp 1–7
- Singh N, Jindal S (2018) Heart disease prediction using classification and feature selection techniques. *Int J Adv Res Ideas Innov Technol* 4(2). [www.IJARIT.com](http://www.IJARIT.com)
- Verma L, Srivastava S, Negi P (2016) A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. *J Med Syst* 40(7):178
- Xue B, Cervante L, Shang L, Zhang M (2012) A particle swarm optimisation based multi-objective filter approach to feature selection for classification. In: Pacific rim international conference on artificial intelligence. Springer, Berlin, pp 673–685
- Zawbaa HM, Emary E, Parv B, Sharawi M (2016) Feature selection approach based on moth-flame optimization algorithm. In: 2016 IEEE Congress on evolutionary computation (CEC). <https://doi.org/10.1109/cec.2016.7744378>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.