IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

A Modified Grey Wolf Optimizer algorithm for feature selection to predict heart diseases

Shaik Khaja Mohiddin[1], Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Susan Peteti[2], **Tummala Swathi[2]**, **Tambura Veera Venkata Harshith[4]**,**Vamshi Krishnamaneni[5]**, **Vatluri Hanusha[6]**
[2,3,4,5,6] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
susanpeteti002@gmail.com[2], swathi.tummala12@gmail.com[3],
harshithram234@gmail.com[4], vamshikrishnamaneni@gmail.com[5],
hanushavatluri57340@gmail.com[6]

DOI:10.48047/IJFANS/V11/I12/180

Abstract

Globally, heart disease is a leading cause of illness and mortality. This impacts people from all around the world. Accurate prediction of the risk of heart disease is crucial for early detection and prevention.For this, large amounts of features/attributes need to be stored and analyzed to diagnose a patient. Storing many features can lead to substandard management of data. We need to store only the chief features. In this study, we proposed a modified grey wolf optimizer for feature selection.

The resultant subset of features is then used to predict the risk of having a heart disease using machine learning model, Support Vector Machine (SVM). We compared the proposed algorithm with the existing GWO-SVM algorithm. We evaluated the effectiveness of the proposed algorithm using accuracy, sensitivity, and specificity metrics. Our results show that, using the modified grey wolf algorithm for feature selection and using SVM weobtained an accuracy of 95.82%, specificity of 94.64%, and sensitivity of 96.86%. The results show the proposed algorithm's capability for predicting the risk of heart disease and could contribute to the development of more accurate and efficient predictive models for heart disease risk assessment.

Keywords: Feature Selection, Grey Wolf Optimizer, Heart Disease Prediction, Machine Learning, Meta-heuristic algorithm, Optimization.

Introduction

With an expected 17.9 million deaths from heart disease in 2019, it is a significant global source of mortality and morbidity (World Health Organization, 2021). Early detection and accurate prediction to the risk of heart disease can help healthcare providers design appropriate interventions to prevent or delay the onset of the disease, as well as reduce the burden of healthcare costs associated with the disease.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

To diagnose a patient with heart disease, data containing various features of the patient need to be stored and analyzed. Storing such large amounts of data can lead to substandard management of data and might even lead to a wrong diagnosis. Therefore, it is important to store only the chief features so that a correct diagnosis can be made while being able to maintain good condition of data. To store only the important data, we reduce the dimensionality of the data. One method of dimensionality reduction is through the use of feature selection algorithms. These techniques can choose the optimal features in a dataset.

There are numerous feature selection methods available, but in recent years, optimization issues have been solved using Nature Inspired Algorithms (NIA). These algorithms are also known a meta-heuristic algorithm because they are used minimize or maximize a solution of a problem.

A novel meta-heuristic algorithm dubbed Grey Wolf Optimizer (GWO) algorithm was released in 2014 by Seyedali*et al.* [1]. The social structure of grey wolves served as an inspiration for this algorithm. The mathematical model for this algorithm is built using their hunting patterns.The two components of the GWO algorithm are exploration and exploitation. The exploration part of the algorithm is the process of exploring the search space. The exploitation part of the algorithm is the process of local searching in promising areas in exploration part.

Hou *et al.*[11] presented a strategy to enhance this search space exploration to determine the path of mobile robots. However, it uses a non-linear convergence factor based on normal distribution for the coefficient 'a', which is not needed for our application. This is taken as inspiration to the algorithm proposed in this paper.

Literature Survey

Many researchers in the past have used Grey Wolf Optimizer algorithm for feature selection as well as for other applications. The social structure and hunting behaviours of grey wolves served as the main inspiration for this algorithm, which was first presented by Seyedali et al. [1] in 2014.

Some of the previous works based on this algorithm where it is specifically used for feature selection in predicting heart disease is using SVM. The GWO approach is used to select features, and an SVM model is then used to predict heart disease using the resulting subset of features [2]. Singh and Kumar proposed various machine learning models that can be

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

used for heart disease prediction [3] whereas Javed *et al.*[4] proposed a different approach by using ensemble learning for heart disease prediction. Eid Emary*et al.*[5] suggested a system to follow when using GWO algorithm integrated in machine learning. It also concludes that GWO performs better than other Nature Inspired Algorithms (NIA).

A new variation of GWO known as RW-GWO that makes use of a random walk to expand the search space was put forth by Preet and Kusum Deep [6].Another NIA for predicting both heart disease and breast cancer was suggested by Sireesha*et al.* [7]. It claims that one of the hybrid models that are suggested outperforms traditional models.

Turabieh [8] proposed a deep learning model ANN is used to predict heart disease. It uses GWO for finding initial weights and biases for inputs to ANN. BakrawyLamiaa shows GWO is integrated with naïve bayes to predict heart diseases. Additionally, the dataset is subjected to a 5-fold cross validation. It states that the results achieved were better than traditional naïve bayes classifier [9].

Finally, Luis Rodr´ıguez*et. al.* shows fuzzy logic i.e., a fuzzy hierarchical operator used in GWO for adding weights to the search agents [10]. These weights are used as an inspiration to the modified algorithm proposed in this paper.

Hou *et. al.* proposed an improved grey wolf optimizer for finding path in mobile robots. This algorithm overcomes the problem of convergence speed and instability in GWO and has a better optimal search capacity [11]. The methodology proposed here is modified to be used for our application. [13-21]

Problem Identification

A. Grey Wolf Optimizer

The population-based optimization algorithm known as the Grey Wolf Optimizer was modelled by the hunting strategies of grey wolves.The social hierarchy of grey wolves serves as the basis for the search agents in this algorithm. Figure 1 shows the social hierarchy of grey wolves.

- Alpha (α): This search agentacts as the best solution. It is responsible for making decision and leading the hunt. After the hunt, the alpha wolf has its share first.
- Beta (β): It search agent acts as the second-best solution. It helps the alpha wolves in hunting the prey.
- Delta (δ): It search agent acts as the third best solution. Delta wolves of the pack are the food providers.

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

- Omega (ω): It represents the solution that follow alpha, beta and delta solutions. It consists of elders or the weakest wolves. They often act as the scapegoat.

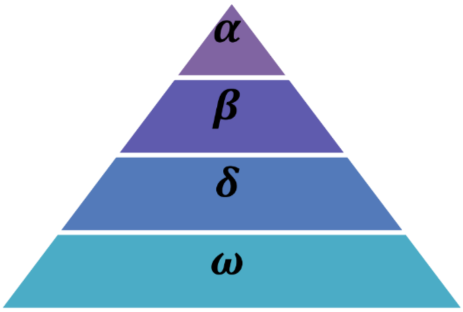

Figure 1: Social hierarchy of grey wolves

B. Mathematical Model of GWO

The hunting process is divided into exploration, encircling and exploitation. The mathematical model is also based on these steps. The encircling of the prey is represented as:

$$D = |C.X_p - X(t)| \quad (1)$$

$$X(t+1) = |X(t) - A.D| \quad (2)$$

$$A = 2.a.r_1 - a \quad (3)$$

$$C = 2.r_2 \quad (4)$$

Where 'a' linearly decreases from 2 to 0 and r1, r2 both are random numbers in the range [0,1].

The parameter 'a' regulates the transition between exploitation and exploration. It is reduced by using the equation from 2 to 0. Eq (5) is shown below:

$$a = 2 - t.\frac{2}{\max\_iter} \quad (5)$$

Where, t represents the current iteration and max_iter represents the maximum number of iterations the GWO algorithm permits.

The exploitation of the prey is represented as:

$$D_\alpha = |\,C.X_\alpha - X(t)|$$

$$D_\beta = \left|\,C.X_\beta - X(t)\right|$$

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

$$D_\delta = | C.X_\delta - X(t) | \quad (6)$$

In terms of alpha, beta, and delta wolves, the respective positions are,

$$X_1 = |X_\alpha - A.D_\alpha|$$

$$X_2 = |X_\beta - A.D_\beta|$$

$$X_3 = | X_\delta - A.D_\delta | \quad (7)$$

The equation to update the position of the search agent is shown below in eq. (8);

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (8)$$

Proposed Methodology

The system architecture of the proposed methodology consists of acquiring data, performing data pre-processing, feature selection and building classification model. Figure 2 shows the system architecture.

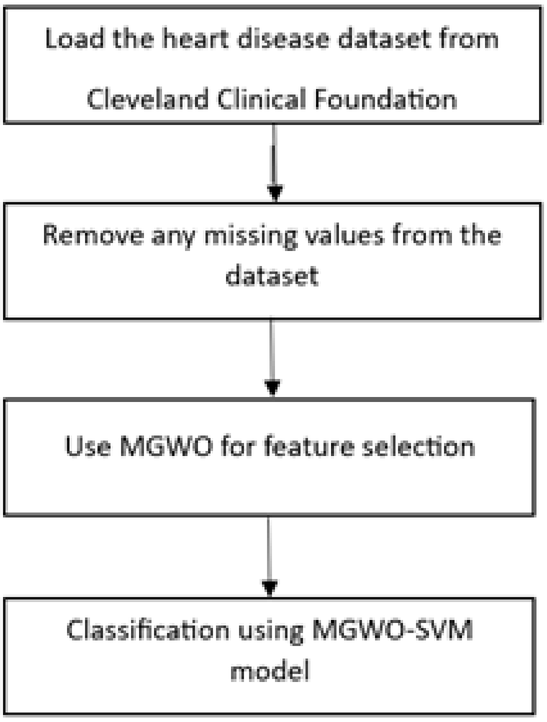

Figure 2: System architecture

Nitin Mittal*et. al.*[12] proposed two ways in which the estimate of parameter a can decrease from 2 to 0. One of those ways is

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

$$a = 2(1 - \frac{t^2}{T^2}) \quad (9)$$

This equation is used in the modified algorithm for parameter a throughout the iterations.

In the exploitation step, the alpha wolf which is the strongest must be closest to the prey to make the hunt successful. The weights are, say, w1, w2, and w3 of alpha, beta, and delta search agents respectively. Hou *et. al.*[11] proposed similar equations for updating the locations of search agents using weights.

Although, this algorithm performs well for the application of finding the path for mobile robots.

However, our application does not use the convergence factor proposed by Hou *et al.* [11]. Using only the equations above mentioned did not give satisfactory results.

This might be because eq. (11) uses the absolute values of X1, X2, and X3 to give importance to the positions.

This is giving values of search agents that are outside the boundaries of the search space because the equation for coefficient 'a' proposed by Hou *et al.* [11] is not used. Therefore, the equations are modified according to the coefficient 'a' that decreases linearly from 2 to 0 (eq. (9)) with the condition that if the positions are still out of bounds of search space they will not be changed.

So, the equations are modified as,

$$W_\alpha = \frac{f_1 + f_2 + f_3}{f_1}$$

$$W_\beta = \frac{f_1 + f_2 + f_3}{f_2}$$

$$W_\delta = \frac{f_1 + f_2 + f_3}{f_3} \quad (10)$$

$$P_1 = \frac{X_1 + X_2 + X_3}{X_1}$$

$$P_2 = \frac{X_1 + X_2 + X_3}{X_2}$$

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

$$P_3 = \frac{X_1 + X_2 + X_3}{X_3} \quad (11)$$

$$X(t+1) = \frac{W_\alpha . P_1 + W_\beta . P_2 + W_\delta . P_3}{W_\alpha + W_\beta + W_\delta} \quad (12)$$

The MGWO algorithm's pseudo-code is shown below:

1.**Initialize**grey wolf population

2.**Initialize** a, A, and C

3. Calculate fitness values of search agents

4. X_α = best solution; X_β = second-best solution; X_δ = third best solution

5.**while** t<max_iter

6.**for** each search agent

7.update a, A, and C

8.calculate the weights and positions using (10) and (11) respectively

9.update the position of current search agent using (12)

10.calculate fitness values of search agents

11.update X_α, X_β, and X_δ.

12.t = t+1

13. return X_α

Implementation

A. Experimental Setup

The MGWO algorithm is developed in MATLAB R2022b and the machine learning models are implemented in Jupyter notebook. They are operated by a computer with an Intel Core i5 processor with 8GB RAM capacity.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

B. Dataset

The dataset is publicly available and one can download it from UCI machine learning repository. The dataset is taken from Cleveland Clinical Foundation.

C. Parameter Setting

The criteria used to set up the suggested grey wolf optimization algorithm are shown in Table 1. The values of lower bound and upper bound changes according to the benchmark function that is being used.

Here, the "F7" benchmark function was used. [1] shows the various benchmark functions that can be used for grey wolf optimization.

Parameters	**Initialized values**
Maximum no. of Iterations	100
No. of Wolves	5
No. of Dimensions	13
Selection Threshold	0
Upper bound	1.28
Lower bound	-1.28

Table 1: Parameter settings for the proposed method

Results and Discussions

The preprocessing and feature selection is performed before building the machine learning models. There were no null values or missing values in the dataset. As mentioned above, Modified Grey Wolf Optimizer is used for performing feature selection and resultant set of features are used for training and testing the models. Then the data is standardized using Standard Scaler.

The models are then created and trained on the resulting set of features. The different machine learning models' performance is determined using measures like accuracy, sensitivity, and specificity.

Table 2 conveys the performance evaluation measures of the proposed algorithm i.e., MGWO-SVM with the existing GWO-SVM. The accuracy of MGWO-SVM is 95.82% and the

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

sensitivity is 96.86% and the specificity is 94.64%. The same set of features are selected with each execution of the algorithm; therefore, the results are the same for each execution.

Method	**Acc.**	**Sens.**	**Spec.**
GWO-SVM	89.83%	93%	91%
MGWO-SVM	95.82%	96.86%	94.64%

Table 2: Comparison of proposed model to existing model

Other machine learning models are also built to make predictions. Their results are shown in Table 3.

Method	**Acc.**	**Sens.**	**Spec.**
MGWO-LR	86.91%	91.62%	81.55%
MGWO-KNN	87.46%	88.48%	86.31%

Table 3: Results of other models

Here, LR stands for Logistic Regression.

The better performance of MGWO compared to existing GWO algorithms might be due to:

- The improved search space exploration using proportional weighting strategy that is inspired from [11]. This helps in finding optimum solution more accurately.
- The dependable way of finding alpha, beta, and delta search agents when implemented in python.

Conclusion

This paper presented a modified Grey Wolf Optimizer algorithm for feature selection to find optimal features in a dataset. The features obtained from MGWO algorithm is then used to build SVM model.

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

It is observed that the model gives an accuracy of 95.82% and sensitivity of 96.86% and a specificity of 94.64%.When loosing even one important feature reduces the accuracy of the model. The proposed algorithm MGWO gives the same set of features with each execution and all the important features are always selected. The performance of MGWO-SVM further supports the conclusion that MGWO is better for the application of feature selection.

Limitations

This study only used one dataset to perform feature selection using MGWO. This study used only a one machine learning model. Accurate predictions also depend on the type of classifier. Working with more advance models might give more accurate predictions.

Future Scope

Larger datasets can be used for feature selection using the proposed algorithm. More types of machine learning models or even deep learning models can be used for predicting heart diseases.

References

[1] SeyedaliMirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, “Grey Wolf Optimizer”, Advances in Engineering Software, Vol. 69, Pages 46-61, March 2014.

[2] Qasem Al-Tashi, Helmi Rais, Said Jadid, “Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification”; In: Advances in Intelligent Systems and Computing; Saeed F, Gazem N, Mohammed F, Busalim A (eds); International Conference of Reliable Information and Communication Technology,Vol. 843, Pages 257-266, September 2018.

[3] Singh A, Kumar R, "Heart Disease Prediction Using Machine Learning Algorithms," In: 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, Pages 452-457, February 2020.

[4] Abdullah Alqahtani, ShtwaiAlsubai, Mohemmed Sha, Lucia Vilcekova, Talha Javed, "Cardiovascular Disease Detection using Ensemble Learning", Computational Intelligence and Neuroscience, Vol. 2022, Article ID 5267498, 9 pages, August 2022.

[5] Emary E, Hossam M Zawbaa, CrinaGrosan, Abul E Hassenian, “Feature Subset Selection Approach by Gray-Wolf Optimization”. In: In: Advances in Intelligent Systems and Computing; Abraham, A., Krömer, P., Snasel, V. (eds); Afro-European Conference for Industrial Advancement, Vol. 334, Pages 1-13, 2015.

[6] Preeti, Kusum Deep, “A random walk Grey wolf optimizer based on dispersion factor for feature selection on chronic disease prediction”, Expert Systems with Applications, Volume 206, November 2022.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[7] SireeshaMoturi, Tirumala Rao S.N., Srikanth Vemuru, "Grey wolf assisted dragonfly-based weighted rule generation for predicting heart disease and breast cancer", Computerized Medical Imaging and Graphics, Volume 91, July 2021.

[8] Turabieh, H, "A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease", American Journal of Operations Research, Vol. 6, Pages 136-146, January 2016.

[9] Lamiaa M El Bakrawy, "Grey Wolf Optimization and Naive Bayes classifier Incorporation for Heart Disease Diagnosis", Australian Journal of Basic and Applied Sciences, Vol. 11, May 2017.

[10] Luis Rodríguez, Oscar Castillo, Jose Soria, Patricia ´ Melin, Fevrier Valdez, Claudia I.Gonzalez, Gabriela E.Martinez, Jesus Soto, "A Fuzzy Hierarchical Operator in the Grey Wolf Optimizer Algorithm", Applied Soft Computing Journal, Vol. 57, Pages 315-328, August 2017.

[11] YuxiangHou, HuanbingGao, ZijianWang, Chuansheng Du, "Improved Grey Wolf Optimization Algorithm and Application", Vol. 22, Issue 10, Page 3810, May 2022.

[12] Nitin Mittal, Urvinder Singh, Balwinder S Sohi, "Modified Grey Wolf Optimizer for Global Engineering Optimization", Applied Computational Intelligence and Soft Computing, Vol. 2016, Article ID 7950348, 16 pages, May 2016.

[13] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[14] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[15] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[16] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[17] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[18] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[19] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[20] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[21] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.