Recursive Journal of Informatics
Vol. 3, No. 1, March 2025

p-ISSN 2963-5551 https://journal.unnes.ac.id/journals/index.php/rji e-ISSN 2986-6588

Optimizing Heart Disease Classification Using the Support Vector Machine Algorithm with Hybrid Particle Swarm and Grey Wolf Optimization

Luthfi Ilham Agus Pratama[1], Anggyi Trisnawan Putra[2]
[1,2]Computer Science Department, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia

Abstract. Heart disease, also known as cardiovascular disease, is a condition that affects the heart and blood vessels, leading to complications such as coronary artery disease, heart failure, arrhythmias, and heart valve disorders. According to the World Health Organization (WHO), approximately 17.9 million people die from heart disease each year. Early detection plays a crucial role in reducing the number of cases and improving patient outcomes.
Purpose: In the era of rapid technological advancements, machine learning has been widely utilized for early diagnosis of heart disease. This study aims to enhance classification accuracy by applying a hybrid PSOGWO (Particle Swarm and Grey Wolf Optimization) method for feature selection and a standard scaler for data balancing in SVM classification.
Methods/Study design/approach: The research begins with obtaining a heart disease dataset, which undergoes preprocessing steps, including feature selection using hybrid PSOGWO and data normalization with a standard scaler. The dataset is then divided into training and testing sets, where the training data is classified using SVM. Performance evaluation is conducted using a confusion matrix to measure accuracy improvements.
Result/Findings: The proposed method successfully selects 10 significant features out of 13 in the dataset. By integrating hybrid PSOGWO with SVM, the classification accuracy improves to 93.66%, representing a 2.44% increase from the original 91.22% obtained using SVM without feature selection.
Novelty/Originality/Value: This research contributes to the development of more effective heart disease prediction models by optimizing feature selection and classification, leading to improved diagnostic accuracy and potential clinical applications.

Keywords: Heart Disease, Classification, SVM, Hybrid PSOGWO, RJI
Received January 01, 2023 / **Revised** May 17, 2024 / **Accepted** March 27, 2025

This work is licensed under a [Creative Commons Attribution 4.0 International License.]

INTRODUCTION

The heart is an organ whose function is to pump blood throughout the body. The heart consists of four main parts which are formed from muscles and electrical impulses as the driving force, which makes the heart function properly, namely the brain and nervous system [1]. If there are abnormalities or damage to blood circulation or blood flow in the cardiovascular system, it will cause complications of heart disease [2]. In 2018 in Indonesia, heart disease was ranked in the top 10 non-communicable diseases. Sufferers are estimated to reach 229,696 people, or around 0.13% of people suffering from heart disease in Indonesia [3]. In Central Java province, heart disease sufferers are estimated to reach 1.6% of the population, or around 3,493 people [4]. There are various ways to help prevent the increase in heart disease sufferers, namely by utilizing the application of classification. Classification of data on heart disease sufferers is very important to help prevent heart disease which continues to increase significantly [5]. Accurate classification results have an important role in carrying out prevention as quickly as possible. Deaths can be minimized when people carry out early diagnosis and enable health workers to provide appropriate treatment and services in a relatively short time [6]. To reach this stage, it is necessary to implement a system that utilizes algorithm features found in machine learning methods as a diagnosis method and utilizes algorithms contained in machine learning methods, one of which is data mining, which of cours e has classification process features [7].

[1]*Corresponding author.
Email addresses: luthfiilhamap@students.unnes.ac.id (Pratama)
DOI: 10.15294/rji.v3i1.737

Machine learning is a part of artificial intelligence (AI) which aims to develop computers to study data and improve their performance without being explicitly programmed [8]. Applying machine learning, there are features that can be utilized well. One application of the method that can be used for classification is data mining, namely the process of developing patterns, information and knowledge sourced from a collection of datasets [9]. In data mining methods, one of the preferred classification methods is the Support Vector Machine (SVM) classification method. The SVM algorithm was first introduced by Vapnik in 1992. SVM is a machine learning algorithm that functions to process classification and regression [10]. In classification, SVM has the aim of finding the best hyperplane to separate data points contained in different classes in a higher dimensional scope [11].

Hybrid PSO is an optimization technique that combines traditional principles of the PSO algorithm with other optimization methods, or in other words, creating problem solving strategies to improve performance and efficiency [12], difficulties in the classification algorithm used and can influence the process of the classification algorithm used. So, a combination of methods is needed to overcome the imbalance of a class, namely by using Grey Wolf Optimization (GWO) [13]. GWO is a metaheuristic optimization algorithm adapted from the social hierarchy and hunting behavior of gray wolves in the wild. The simulation of this algorithm is a leadership hierarchy and collaborative hunting strategy for wolves to obtain optimal solutions and iteratively in a complex optimization [14]. The main contribution of this is the development of a combined SVM Hybrid PSOGWO method to help prevent and predict heart disease [15]. Hybrid PSO modified with GWO and supported by SVM is used to select the best features. Hybrid PSO is used to target rejected individuals from each generation to fulfill the concept of recovering rejected individuals, maximizing the use of all individuals from each generation [16].

In previous research to classify high and low size data on diabetes diagnosis in pregnant women using fuzzy rule optimization and combined with GWO, then compared with Ant Colony Optimization (ACO) which has been combined with fuzzy rules [17]. GWO can achieve superior results compared to ACO. GWO can produce statistically significant data increments, ultimately having advantages over other optimization algorithms. With its nature of imitating wolf hunting, GWO can of course display and produce better data, producing an accuracy of 81%, so that in the future GWO can be applied to real world applications with higher sizes and can be used with general classification algorithms such as SVM, PSO, and etc [18]. In this research, a classification process will be applied using the SVM algorithm combined with hybrid PSOGWO with a dataset sourced from a heart disease dataset obtained from the open-source UCI Machine Learning dataset. Classification applied with the aim of finding out whether an algorithm has better performance when the algorithm is applied.

METHODS

The process that will be carried out in this research is implementing the feature and class selection algorithm balancing algorithm as an optimization algorithm to understand whether the applied algorithm has significant influence or not. The framework in this research is to compare the accuracy of a classification algorithm without feature selection and a classification algorithm that uses feature selection (hybrid PSOGWO) by applying an algorithm to classify heart disease datasets. The framework of the model that will be carried out can be seen in Figure 1.

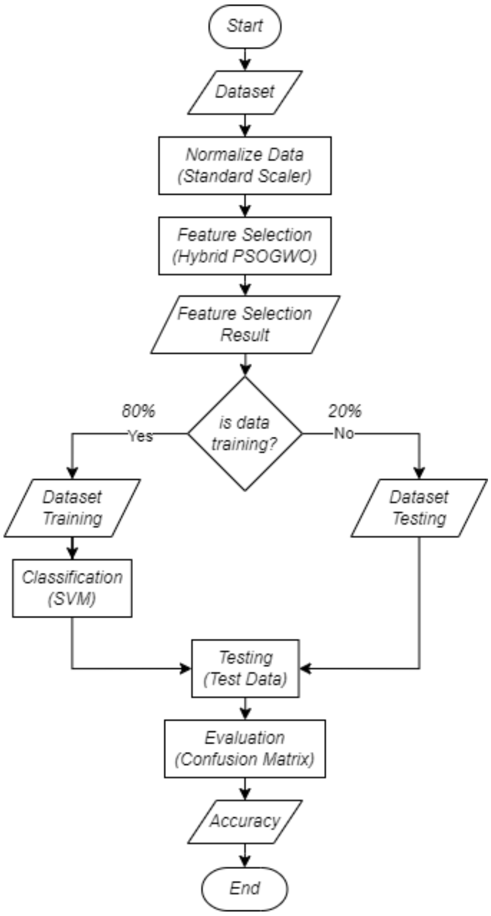

Figure 1. Flowchart research design

RESULT AND DISCUSSION

In this study, a heart disease dataset obtained through UCI was used. It consists of 1025 lines of data and 13 columns showing the condition of each patient. In this data set, the label column contains numeric data with the value '1' indicating that the patient has suffered from heart disease and '0' indicates that the patient does not suffer from heart disease. In this data set, the heart disease dataset has 1025 instances with 526 instances of class "1" (positive) and 499 instances of class "0" (negative). An unbalanced data ratio between classes indicates that the dataset used is unbalanced. This data set can be accessed via the link https://archive.ics.uci.edu/dataset/45/heart+disease. An example from a dataset is shown in Table 1.

Table 1. Sample data from heart disease dataset

No	Age	Sex	…	Thal	Target
1	52	1	…	3	0
2	53	1	…	3	0
3	70	1	…	3	0
4	61	1	…	3	0
5	62	0	…	2	0
…	…	…	…	…	…
1021	59	1	…	2	1
1022	60	1	…	3	0
1023	47	1	…	2	0
1024	50	0	…	2	1
1025	54	1	…	3	0

Experimental Study

In this research, experiments were carried out using open-source datasets obtained publicly. That The data set was then processed on a local computer running on these specs: Intel i7-7700HQ (2.80GHz base), 8 GB DDR4 ram, and 240 GB SSD + 1 TB HDD storage. This program runs in Visual Studio The code runs on Windows 10. Almost all algorithms and applications are written using Python programming language running Python 3.9.15. The library used in this research is: numpy to do complex computations on data, pandas for processing datasets, scikit-learn used for preprocessing, divide the data into training and test data, and to show the classification accuracy and perplexity matrix, imbalance-learning is used to resolve imbalances in data sets using edited nearest neighbors, and for classification methods using SVC from SVM.

Data Preprocessing

At this stage, the data is processed through several processes so that the data is easier to apply to classification process. The steps applied at this stage are missing value treatment, data normalization, and data coding. Missing value treatment is carried out by using the mean value to fill in valueless data. Data Normalization is carried out using a standard scaler by scaling the previously distributed data to a normal scale from 0 to 1. Finally, data coding is applied using a label encoder that transforms categorical data to numerical data based on category. Data after the preprocessing stage is presented in Table 2.

Table 2. Sample data from dataset after preprocessing

No	Age	Sex	…	Thal	Target
1	0,2684	1	…	3	0
2	0,1582	1	…	3	0
3	1,7166	1	…	3	0
4	0,7241	1	…	3	0
5	0,8344	0	…	2	0
…	…	…	…	…	…
1021	0,5035	1	…	2	1
1022	0,6138	1	…	3	0
1023	0,8198	1	…	2	0
1024	0,489	0	…	2	1
1025	0,0479	1	…	3	0

Feature Selection Using Hybrid PSOGWO

The feature selection process is carried out using hybrid PSOGWO, which targets rejected individuals from each generation to fulfill the concept of recovering rejected individuals, maximizing the use of all individuals from each generation. After the feature selection process is applied to the cleaned dataset, it produces a feature set that has the greatest fitness value in the population created. There are 10 of the 13 features selected in this process, as shown in Table 3 and Figure 2.

Table 3. Selected feature after hybrid PSOGWO

No.	Feature Name
1	Age
2	Sex
3	Cp
4	Trestbps
5	FBS
6	Restecg
7	Exang
8	Slope
9	Ca
10	Thal

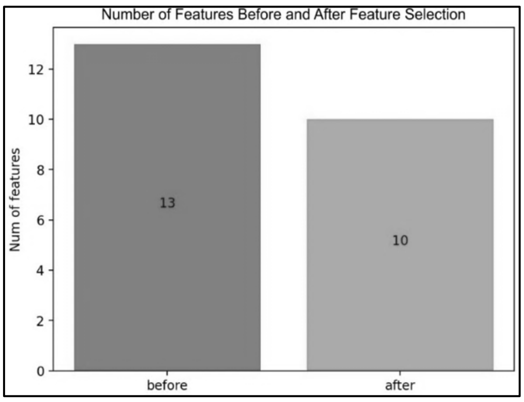

Figure 2. Comparison of the data before and after feature selection

Split Data

Split data is carried out using a ratio of 80:20, namely 80% for training data which will go through the classification process using SVM, and 20% for testing which will later be evaluated, the explanation can be seen in Figure 3, Figure 4, and Figure 5.

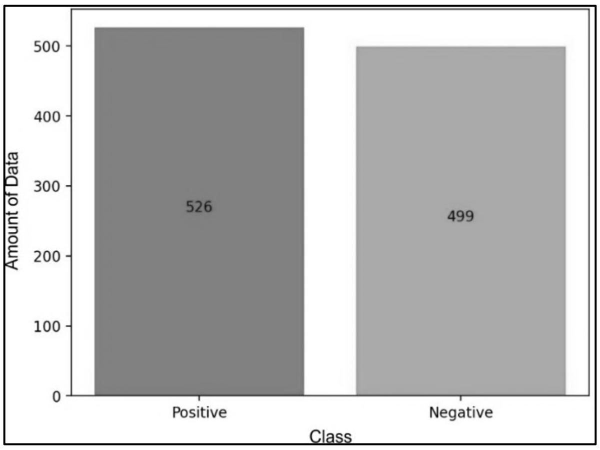

Figure 3. The whole dataset

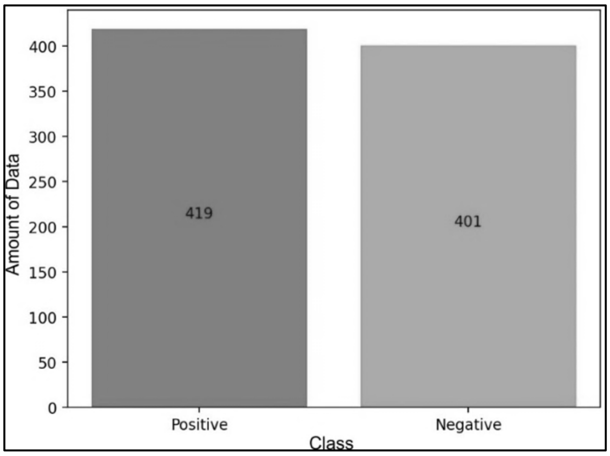

Figure 4. Data training

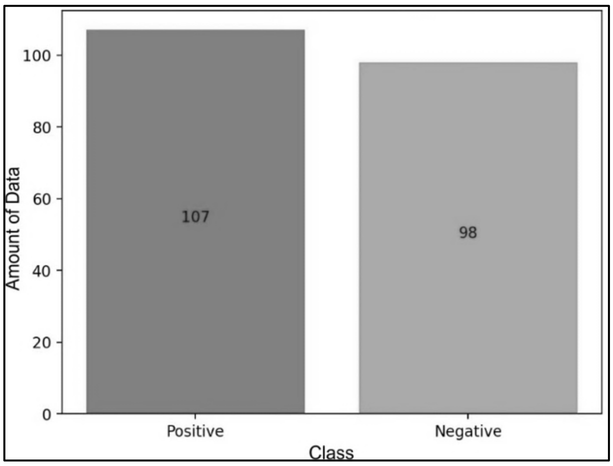

Figure 5. Data testing

SVM Classifier

In this study, classification was carried out using the SVM method. Comparison of accuracy results was carried out by comparing the accuracy results of classification using the SVM algorithm without using feature selection from hybrid PSOGWO and compared with the classification accuracy results of the SVM algorithm combined with feature selection using hybrid PSOGWO. The explanation can be seen in Table 4 and Table 5.

Table 4. Classification report SVM

	Precision	Recall	f1-score	Support
Negative	95,45	85,71	90,32	98
Positive	88,03	96,26	91,96	107
Macro AVG	91,74	90,99	91,14	205
Weighted AVG	91,58	91,22	91,18	205

Table 5. Classification report SVM + hybrid PSOGWO

	Precision	Recall	f1-score	Support
Negative	93,81	92,96	93,33	98
Positive	93,52	94,39	93,95	107
Macro AVG	93,67	93,62	93,64	205
Weighted AVG	93,66	93,66	93,66	205

Comparasion Accuracy

Accuracy comparisons have been produced from the process that has been passed, namely the classification accuracy results of the SVM algorithm without using feature selection from the hybrid PSOGWO, which is 91.22%, while the classification accuracy results of the SVM algorithm combined with feature selection from the hybrid PSOGWO are 93. 66%. It was proven that there was a significant increase of 2.44% resulting from a combination of methods between SVM and hybrid PSOGWO. A description of the evaluation is in Figure 6 and Figure 7.

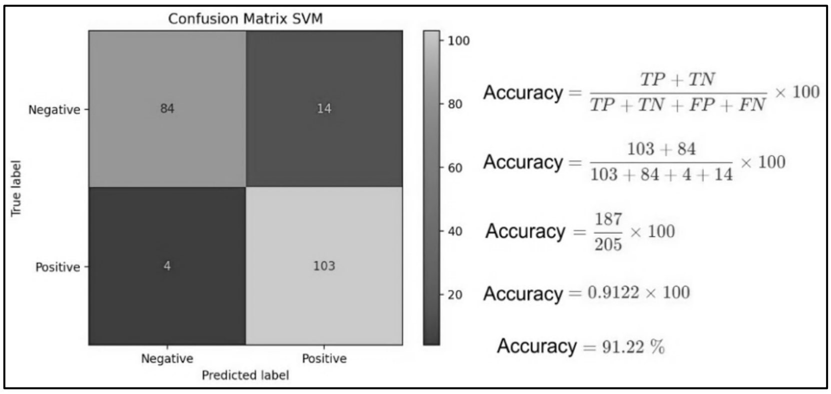

Figure 6. Confusion matrix SVM

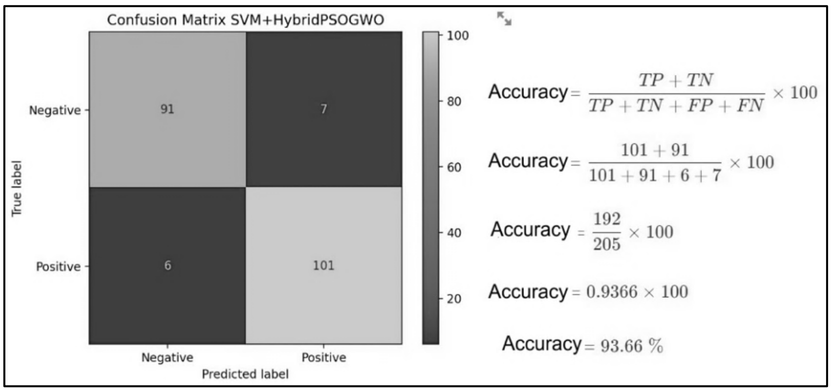

Figure 7. Confusion matrix SVM + hybrid PSOGWO

Comparison Of the Proposed Method and Related Works

During this research, there were many related works that inspired and helped complete this experiment. The references used are works that use similar methods and data which also discuss data mining, classification, and using heart disease datasets. In this section, we compare the results from references using the same dataset and the proposed method to get an idea of the significance of the experiments performed. Comparison between the proposed method and the reference method which uses the same dataset as shown in Table 6.

Table 6. Comparison of related works result and proposed method

No.	Author	Dataset	Method	Accuracy
1	Kumar & Kumar (2021) [2]	Heart Disease	ANN & IMGWO	91,43%
2	Sugendran & Sujatha (2023) [1]	Heart Disease	Enhanced Genetic, Fuzzy Weight updating SVM & K-Means Clustering	91,68%
3	Al Tashi et al. (2019) [14]	18 Heart Disease Dataset	Hybrid PSOGWO & KNN	93%
4	Cherian et al. (2020) [3]	Heart Disease	Lion Algorithm update PSO & Neural Network	87,41%
5	Shah et al. (2020) [4]	Heart Disease	MFFSA, AFSA & SVM	84,2%
6	Proposed Method	Heart Disease	SVM & Hybrid PSOGWO	93,66%

CONCLUSION

The results of the application in this research are using a standard scale to equalize the data so that there is no bias between columns in the dataset. Datasets that have gone through the data equalization process will have the best features selected using the PSOGWO hybrid algorithm method. Hybrid PSOGWO produces the best number of features, where in the initial dataset there are 13 features, then after going through the feature selection process, the best features become 10 features. The dataset that has gone through the preprocessing process is then divided into 2 parts, namely training data and testing data with a ratio of 80:20. Then the next process on the training data will use the classification process of the SVM algorithm. Next, the model results are tested using testing data, then the classification results will be evaluated using a confusion matrix. Thus, the application of the data equalization method from the standard scaler and feature selection from the hybrid PSOGWO can be used as an optimization method for SVM. The accuracy results of the heart disease dataset classification using the SVM algorithm without combining it with the hybrid PSOGWO was 91.22%. 3. The accuracy results of the SVM algorithm classification combined with hybrid PSOGWO reached 93.66%. Thus, it can be concluded that the implementation of hybrid PSOGWO can increase the accuracy resulting from the SVM algorithm, namely by 2.44% in the classification of heart disease.

REFERENCES

[1] G. Sugendran and S. Sujatha, "Earlier identification of heart disease using enhanced genetic algorithm and fuzzy weight based support vector machine algorithm," *Measurement: Sensors*, vol. 28, no. 4, pp. 1-2, 2 June 2023, doi: 10.1016/j.measen.2023.100814.

[2] N. Kumar and D. Kumar, "IMGWO based ANN: A new heart disease diagnosis model to classify real world dataset," *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 12, no. 11, pp. 1001-1002, July 2021, doi: 10.21817/indjcse/2021/v12i4/211204182.

[3] R. P. Cherian, N. Thomas and S. Venkitachalam, "Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm," *Journal of Biomedical Informatics*, vol. 110, no. 9, pp. 1-2, October 2020, doi: 10.1016/j.jbi.2020.103543.

[4] S. M. S. Shah, F. A. Shah, S. A. Hussain and S. Batool, "Support Vector Machines-based Heart Disease Diagnosis using Feature Subset, Wrapping Selection and Extraction Methods," *Computers & Electrical Engineering*, vol. 84, no. 3, p. 2, 5 March 2020, doi: 10.1016/j.compeleceng.2020.106628 004.

[5] S. Nashif, M. R. Raihan, M. R. Islam and M. H. Iman, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," *World Journal of Engineering and Technology*, vol. 6, no. 4, pp. 854-873, 22 November 2018, doi: 10.4236/wjet.2018.64057.

[6] A. Mustaqeem, S. M. Anwar, A. R. Khan and M. Majid, "A statistical analysis based recommender model for heart disease patients," *International Journal of Medical Informatics*, vol. 108, no. 2, pp. 134-145, 12 October 2017, doi: 10.1016/j.ijmedinf.2017.10.008.

[7] H. Turabieh, "A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease," *American Journal of Operations Research*, vol. 6, no. 5, pp. 136-146, 18 March 2016, doi: 10.4236/ajor.2016.62016.

[8] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems With Applications*, vol. 166, no. 4, pp. 1-22, 24 September 2020, doi: 10.1016/j.eswa.2020.114060.

[9] N. Jothi, N. A. Rashid and W. Husain, "Data Mining in Healthcare – A Review," *Procedia Computer Science*, vol. 72, no. 11, pp. 306-313, 23 December 2015, doi: 10.1016/j.procs.2015.12.145.

[10] D. A. Anggoro and N. D. Kurnia, "Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 1689-1693, May 2020, doi: 10.30534/ijeter/2020/32852020.

[11] J. Cervantes, F. G. Lamont, L. R. Mazahua and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, no. 5, pp. 189-215, 30 September 2020, doi: 10.1016/j.neucom.2019.10.118.

[12] T. Joachims, "Svmlight: Support vector machine," *Cornell*, vol. 6, no. 8, pp. 1-13, 14 August 2008, doi: 10.1145/1007730.1007733.

[13] S. S. G. and M. K., "Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization," *Pattern Recognition Letters*, vol. 125, no. 3, pp. 432-438, 5 June 2019, doi: 10.1016/j.patrec.2019.06.005.

[14] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili and H. Alhussian, "Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection," *Digital Object Identifier*, vol. 7, no. 9, pp. 39496-39508, 21 March 2019, doi: 10.1109/ACCESS.2019.2906757.

[15] H. Faris, I. Aljarah, M. A. Al-Betar and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," *Neural Computing and Applications*, vol. 30, no. 12, pp. 413-435, 13 November 2017, doi: 10.1007/s00521-017-3272-5.

[16] H. Rezaei, O. Bozorg-Haddad and X. Chu, "Grey Wolf Optimization (GWO) Algorithm," *Advanced Optimization by Nature-Inspired Algorithms*, vol. 720, no. 10, pp. 81-91, 1 July 2017, doi: 10.1007/978-981-10-5221-7_9.

[17] K. Polat and S. Gunes, "A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS," *Computer methods and programs in biomedicine*, vol. 88, no. 2, pp. 164-174, 31 July 2007, doi: 10.1016/j.cmpb.2007.07.013.