International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 12 ǁ December 2022 ǁ PP. 599-606

An Ensemble Technique for Heart and Liver Disease Prediction using Random Walk Grey Wolf Optimization

Ashima, Research Scholar
Swami Vivekanand Subharti University,Meerut
ashimamunja2310@gmail.com

Dr. Manoj Kapil, Professor
Swami Vivekanand Subharti University,Meerut
kapil.manoj@gmail.com

Abstract*- Nowadays, machine learning is commonly used in a number of industries. Since, there is an abundance of data readily available, machine learning is used as an effective assisting mechanism in clinical diagnostics. As a result of heavy alcohol consumption, inhalation of toxic gas, narcotics, food contamination, and an unhealthy lifestyle, people with liver and heart disease are becoming more prevalent. Over the world, both liver and cardiac diseases have a high death rate. To save lives, it is essential to identify these disorders as soon as possible. Health care organizations that implement machine learning classification algorithms see amazing results that improve the efficiency and accuracy of disease diagnosis. Tools and techniques for machine learning help to extract usable data from datasets, producing more precise results. In this study, aRandom Walk grey wolf optimization model with the adaptive boosting (AB) approach is used to develop a hybrid model for the categorization of heart and liver data. The machine learning repository at UCI is where the data sets are sourced from. The results are computed based on classification accuracy, error, correctness, recall, and F1 score. The results are compared to the Random Walk grey wolf optimization adaptive boosting algorithm.(RWGWOAB).*
Keywords*: Heart Disease, Machine Learning, Liver Disease, Adaptive Boosting, Grey wolf optimization algorithm*

I. Introduction

Machine learning is a branch of computer science that aims to improve computer innovation. There are several applications of machine learning in everyday life, particularly in the realm of healthcare. Machine learning has many components, including feature extraction, feature selection, algorithm selection, training, and testing. Due to its robust data analysis skills, machine learning is crucial in the healthcare industry. Typically, machine learning techniques which reduce diagnosis time, increase accuracy and efficiency are used by scientists to convey their interest in prediction and diagnosis. Any type of disease can be identified using supervised machine learning algorithms. Although this study will concentrate on detecting heart and liver diseases, supervised machine learning algorithms can be used to diagnose any type of disease. In order to save lives, heart illness must be correctly diagnosed quickly or appropriately diagnosed if it is heart disease. Heart and liver disease are increasingly recognized as one of the top causes of death worldwide. The liver and the heart are regarded as the two most important tissues in the body of an individual. As a result, cardiac and liver diseases are frequently considered to be serious health issues. Heart and liver diseases are the frequent causes of unexpected mortality in affluent nations, according to multiple surveys. Infants are now affected by liver and heart diseases. As a result, monitoring for liver and cardiac diseases is frequently done in daily life [1].

To improve data classification accuracy, numerous researchers in the medical area have tried with a variety of techniques. Better accurate classification methods will provide additional data that can be used to identify potential patients and improve the accuracy of diagnoses. Many optimization approaches, including GWO, the Firefly algorithm (FA), and GA, as well as machine learning classifiers such as decision tree (DT), random forest (RF), K- nearest neighbour (KNN), and logistic regression (LR), have been utilized to predict heart and liver disease. [2] offers a novel system for classifying cardiac arrhythmia beats. The Pan-Tompkins algorithm for R-peak recognition, the discrete orthogonal Stockwell transform (DOST) for extracting features from ECG signals, Adaptive Boosting (AB) for classification for automatic cardiac arrhythmia beat classification, and the Grey Wolf optimization (GWO) technique for parameter modification are all used in this method. In [3] used Nave Bayes and Adaptive Boosting classification algorithms to identify liver diseases. The precision of classification and execution time are the performance parameters that are utilized for evaluating these classifier algorithms. The trials' findings indicate that the AB is a superior classifier for identifying liver diseases. In [4], paper main concept is to employ several classification methods to forecast liver disease. The algorithms used in this research are Adaptive Boosting, K-Nearest Neighbor, and Logistic Regression. We conclude that liver disease can be accurately predicted by logistic regression. In [5], researchers explore the efficacy of KNN, decision trees, linear regression, and Adaptive Boosting, among other machine learning algorithms, for predicting cardiac illness by training and testing on the UCI repository dataset. In [6], opinion mining applies feature selection approaches based on decision trees to predict cardiac events using fewer variables and more precision. In [7], it is suggested that unsupervised rough set techniques be used for the task of text grouping web opinions. The classification accuracy of the current methods for liver diseases is, however, still subpar and minor enough to be applied in real-world settings. Strong diagnosis is required because liver and heart disorders are becoming more severe and because the death rate can be reduced with effective treatment. The categorization accuracy of current methods for liver and cardiac conditions is still subpar and minor enough to be ignored in real-world settings. In this research, we aim to develop a hybrid model for predicting cardiovascular and hepatic diseases utilizing a variant of the Grey Wolf Optimization technique and Adaptive Boosting. In this research, we aim to develop a hybrid model for predicting cardiovascular and hepatic diseases utilizing a variant of the Grey Wolf Optimization technique and Adaptive Boosting. The current machine learning classifiers and hybrid algorithms cannot accurately forecast illnesses. We developed a hybrid model to enhance disease prediction accuracy and address the method's limitations by merging the enhanced GWO approach with AB. The remaining sections of the research are organized as follows. The study that is linked to it is explained in Section 2. The proposed methodology is spelt out in detail in Section 3. The supporting data and conclusions are included in Section 4. After that, Section 5 present the conclusion.

II. Literature Review

There have been numerous recent efforts to improve the GWO algorithm in various ways. Some variants suggested methods for modifying the GWO's a and c settings. Other efforts have tried to improve GWO by adding new or different operators, such as local search methods. Finally, by combining GWO with other metaheuristic algorithms (hybrid algorithms), we can enhance GWO's qualities like its exploration/exploitation balance. Related studies also focus on the topic of adaptive adjustment of the parameters. Adaptive techniques have been utilized in tandem with other metaheuristic algorithms before, and the results were encouraging. Our proposed adaptive algorithm, which is fitness-dependent, is fundamentally different from the existing methods. However, we delve into the fundamentals and characteristics of a wide range of adaptive algorithms. GWO can be optimized for improved performance on particular types of problems by modifying its updating scheme or other factors. In place of a straightforward average, wdGWO [8] considers the spread of the top three choices. The exploration is improved in mGWO proposed by Mittal et al. [9] by adjusting the parameters of convergence to nonlinear form. To achieve a better equilibrium in limited optimization problems, the modified augmented Lagrangian with improved grey wolf optimizer (MAL-IGWO) [10] uses a nonlinear adjustment for the exploration parameter of the GWO. To improve exploration over exploitation, another modified GWO method is optimized for searching small populations by adjusting the parameter a. However, the algorithm's performance is enhanced by these nonlinear tweaks for some classes of situations. Exploratory favourable modification of mGWO improves convergence performance on unimodal functions but fails on more complex multimodal functions. In addition to the parameter update equations [11,12], fuzzy logic is used for the dynamic adaptation of GWO parameters and the update rule for the position of agents. To enhance solutions to the partial discharge optimization problem [13], the parameter an is adaptively changed based on the relative function values of the population. However, fitness values are used at every iteration, and like previous variants, this method is highly dependent on the number of iterations. Modifications are made to both the update step and the nonlinear parameter setting of the exploration/exploitation parameter an in EE-GWO [14], making it an exploration-enhanced GWO algorithm. It is demonstrated that high-dimensional complicated issues, as opposed to simpler unimodal functions, are better suited to EE-GWO. Instead of changing parameter A, Random Opposition Based Learning GWO [15] increases exploration to enhance the algorithm. Using opposition-based learning to improve variety, Enhanced GWO (EGWO) then adjusts the parameter to fluctuate throughout the initial half of the optimization before stabilizing at a constant value during the remaining iterations [16]. The author also employed supervised machine learning in another research [17] to automatically identify white areas in liver biopsies. The authors of the research state that a supervised machine learning classifier was used to examine liver images with the help of annotations provided by two human pathologists. Bayesian Classification is employed by the author in order to forecast liver disease [18]. One of the most common classification models, Bayesian classification was utilized in that study. Hepatitis is an infectious disease that attacks the liver. Artificial Neural Networks were utilized in a study for the detection of hepatitis-related liver illness. [19]. We also used previous studies' results on hepatitis diagnosis as a basis for our own. A model for extracting data from EHRs has been created by AakashChauhan et al. using association rule generation and ML association mining. The model is helpful for understanding the overall picture and underlying trends in a patient's data [20]. Saqlain et al. [21] built a model with the Fisher score method for feature selection and an SVM classifier for the prediction model, and it achieved an accuracy of 81.91 percent, a sensitivity of 72.92 percent, and a specificity of 88.68 percent. Latha and Jeeva [22] constructed a hybrid framework with an accuracy of 85.48 percent using a variety of ensemble learning techniques and four distinct machine learning (ML) classification algorithms (NB, BN, RF, and MP).

III. Proposed methodology

The Random Walk Grey Wolf Optimizer based (RWGWO) is a metaheuristic optimization method inspired by grey wolf hunting behaviour in the wild. It is an evolution of the Grey Wolf Optimizer (GWO), a nature-inspired optimization algorithm proposed by Mirjalili et al. in 2014.

In RWGWO, the algorithm incorporates a random walk strategy along with the standard GWO mechanism. This random walk introduces a stochastic element, which can help the algorithm escape local optima and explore the search space more effectively.The architecture of proposed work is represented by figure 1.

3.1 RWGWO Algorithm

***Step 1**. Initialization*

a. Initialize a population of grey wolves (solutions).

b. Define the fitness-function to evaluate the quality of solutions.

***Step 2.** Grey Wolf Pack Hierarchy*

a. The grey wolves are divided into four categories: Alpha, Beta, Delta, and Omega. These represent the leader positions in the search space.

b. The alpha wolf has the best fitness, followed by beta, delta, and omega.

***Step 3.** Random Walk Factor*

a. Introduce a dispersion factor (DF) which controls the extent of the random walk behavior. This factor influences how much exploration is done in the search space.

***Step 4.**Iteration:*

Perform the following steps for a specified number of iterations or until convergence criteria are met:

***a)** Update Grey-Wolf Positions:*

Update the position of each grey wolf based on its current position, the leader positions, and the random walk factor.This step includes both the standard GWO movement and the random walk component.

***b)** Evaluate Fitness:*

Evaluate the fitness of each wolf depend on the objective function.

***c)** Update Pack Hierarchy:*

Update the positions of alpha, beta, delta, and omega wolves depend on their fitness-values.

***d)** Apply Random Walk:*

Apply the random walk component to each wolf's position based on the dispersion factor.

***e)** Update Dispersion Factor:*

Optionally, update the dispersion factor based on some adaptive strategy.

***f)** Convergence Check:*

Check for convergence based on a predefined criterion.

***Step 5**. Return Best Solution:*

a. Once the algorithm converges or a stopping criterion is met, return the best solution found.

The dispersion factor (DF) in RWGWO plays a crucial role in the balancing exploration and exploitation. A higher DF leads to more exploration, potentially helping the algorithm escape local optima. However, a very high DF may result in excessive exploration and slow convergence.

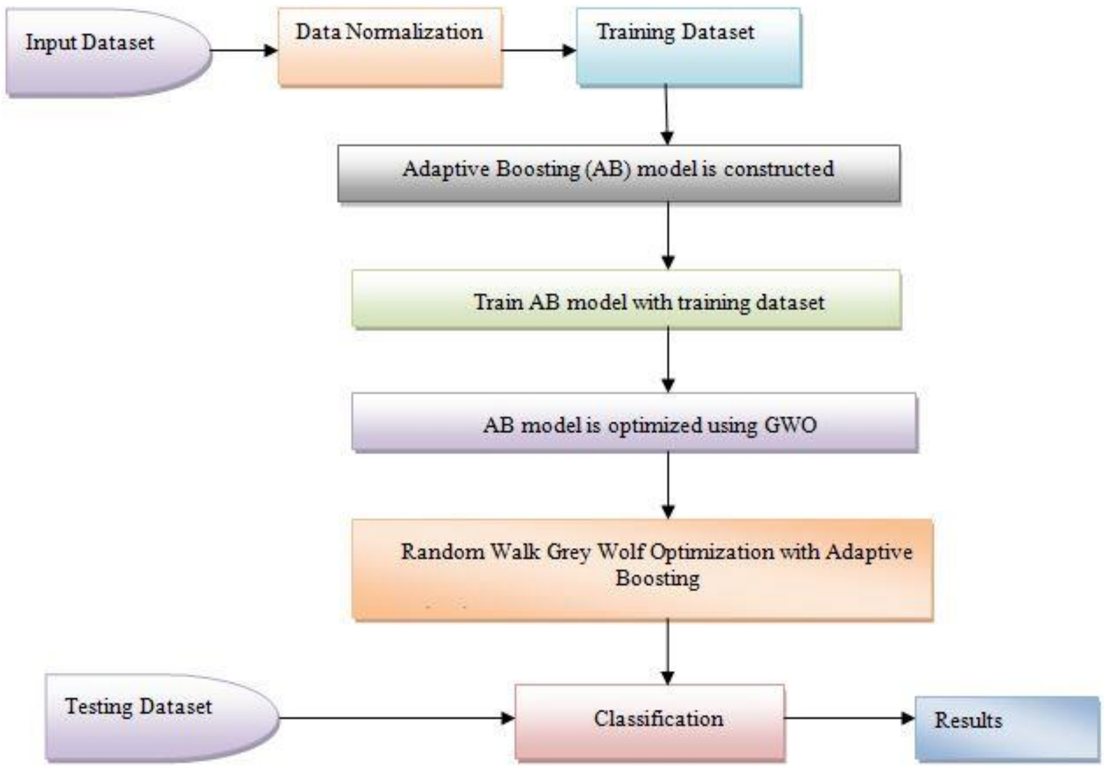

Figure 1. Proposed work Architecture

IV. Result Analysis

In this part, we provide the results of the RWGWOAB analysis of the heart and liver dataset using the precision, accuracy, error, recall, and F1 score. In the experiment, the dataset is split into two parts: the training set and the test set. This ratio accounts for 80% of the data in the training set. The experiment is coded in Python and utilizes the libraries pandas, sci-kit Learn.

4.1 Performance Measures

1.1.1. Accuracy: A standard measure of a classification algorithm's efficacy. The score is based on the ratio of accurate predictions to total accurate forecasts.

1.1.2. Classification Error: Classification error refers to a type of measurement mistake that occurs when respondents give a false answer to a survey question.

1.1.3. Precision: It is the number of data that the machine learning algorithm correctly created and is utilized in data retrieval.

1.1.4. Recall: The recall of machine learning model is the frequency with which it produces a positive result.

1.1.5. F1-score: The estimated value comes from averaging the weighted measures of accuracy and reliability.

The entiremachine learning method explorations in this study was carried out using the Scikitlearn library and the Python programming language. The AB classifier is optimized using random walk grey wolf optimization, and the AB classifier is trained with 200 iterations. The primary objective of this study is to identify the optimal categorization outcomes for the heart and liver illness dataset. The 0.8 databases can be easily divided into a 0.2 testing set following a simple rule of thumb.

Table 1. Parameter used in the proposed RWGWOAB.

Sr. No.	Parameters Used	RWGWOAB
1	**W**	0.6
2	**C2**	0.9
3	**Pr**	1
4	**N**	1
5	**Target Error**	1
6	**C1**	0.5

7	**Vdcraziness**	0.95
8	**Sgnr**	-1
9	**Iteration**	200
10	**Particles in number**	20

The experimental parameters for the proposed RWGWOAB algorithm's testing on the heart and liver datasets are presented in Table 1. The values and parameters utilized are shown in this table. C1 = 0.5, Vdcraziness = 0.95, Sgnr = -1, Iteration = 200, Number of Particles = 20, and the Target Error = 1.

The classification outcomes for AB, GWOAB, and RWGWOAB on the heart dataset are shown in Table 2 in that order. The precision, recall, and F1 score for each classification method for heart disease prediction outlined above are shown in Table 2. With 97.14% recall, 91.89% precision, and 94.44% .F1 score for the heart dataset, proposed algorithm RWGWOAB outperforms other algorithms, as shown in Table 2. Figure 2 is a graphical representation of results of heart disease classifications.

TABLE 2. Heart Dataset Classification

Techniques Used	**Precision**	**F1 Score**	**Recall**
AB	90.91	89.55	88.24
GWOAB	93.94	91.17	88.57
RWGWOAB	91.89	94.44	97.14

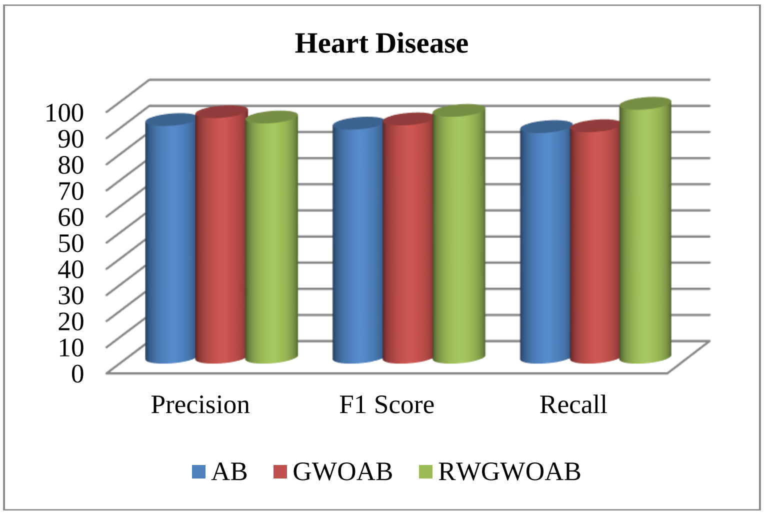

Figure 2. Classification for Heart Dataset

The classification outcomes for the liver dataset using SAB, GWOAB, and RWGWOAB are shown in Table 3. The precision, recall, and F1 score for each classification method for the prediction of liver disease are shown in Table 3. For liver datasets, Table 3 provides 100% precision, 97.41% recall, and 98.68% F1 score. The result is presented by figure 3.

TABLE 3. Liver Dataset Classification

Techniques Used	**Precision**	**F1 Score**	**Recall**
AB	100	77.24	62.93
GWOAB	100	91.07	83.62
RWGWOAB	100	98.68	97.41

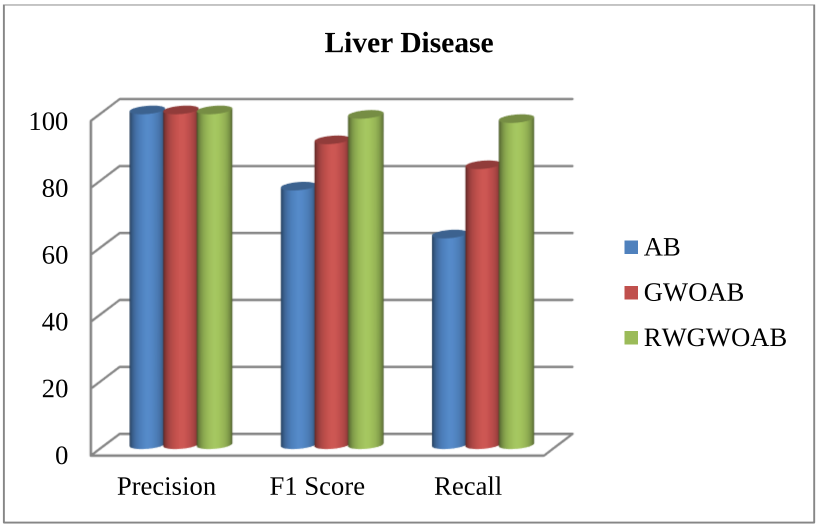

Figure 3. Classification for Liver Dataset

On the heart dataset, Table 4 displays the best accuracy each algorithm achieved using the best ideal hyper-parameters. On the heart dataset, the suggested RWGWOAB had an error rate of 07.41% and 92.59% accuracy according to Table 4. The accuracy and error results are presented in figure 3.

TABLE 4.Accuracy and Error of classifiers to Heart dataset

Techniques Used	**Accuracy (%)**	**Error (%)**
AB	87.04	12.96
GWOAB	88.89	11.11
RWGWOAB	92.59	7.41

The best accuracy generated by each method using the best liver dataset with optimal hyper-parameters is shown in Table 5. The suggested RWGWOAB for the liver dataset receives 97.41% accuracy with a 02.59% error rate from Table 5. The accuracy and error results are presented in figure 4 for liver.

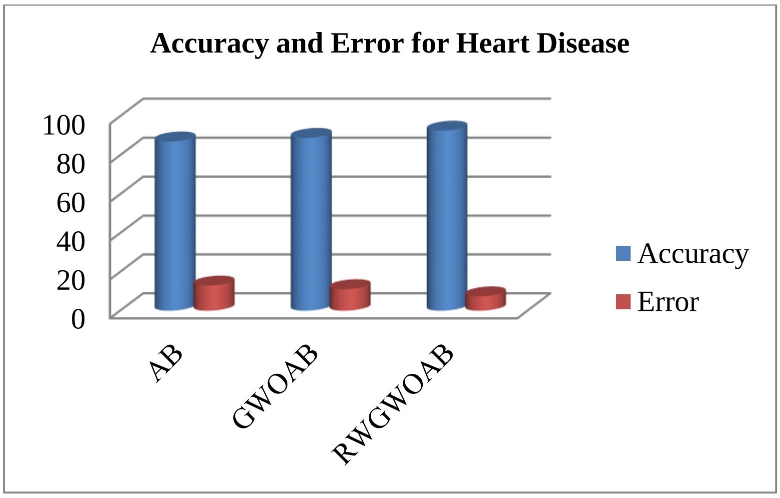

Figure 3. Accuracy and Error for Heart Dataset

TABLE 5.Accuracy and Error of classifiers to Liver dataset

Techniques Used	**Accuracy (%)**	**Error (%)**
AB	62.93	37.07

GWOAB	83.62	16.38
RWGWOAB	97.41	2.59

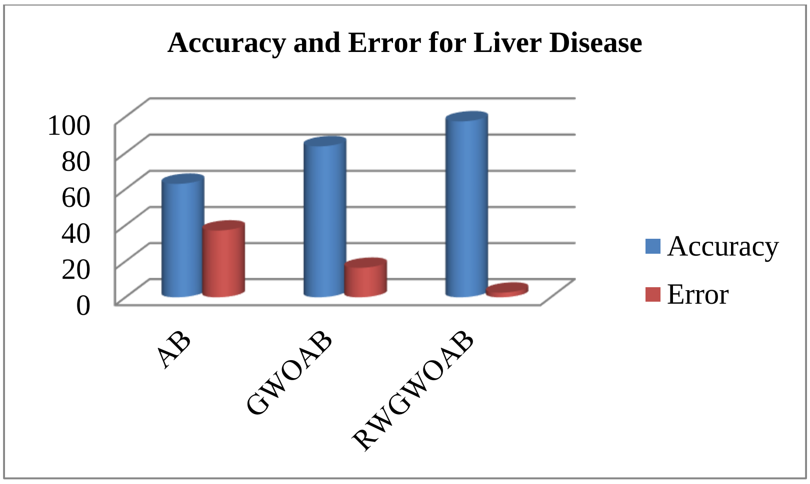

Figure 3. Accuracy and Error for Liver Dataset

V. Conclusion

The RWGWOAB is used for predicting liver and heart disorders. The confusion matrix, classification accuracy, classification error rate, precision, recall, and F1 score have all been computed and examined to determine each algorithm's performance. The proper experimental analysis leads to the conclusion that the intended RWGWOAB provides superior classification results, including the greatest classification rate and the lowest error rate, when used to the prediction of heart and liver disorders. Although only the heart and liver have been employed in the research so far, the disclosed hybrid algorithm may one day be used for the prediction of other diseases as well. Further research will involve applying our suggested hybrid algorithm with additional parameters to larger data sets covering more diseases.

References

[1]. Vijayashree, J., & Sultana, H. P. (2018). A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. Programming and Computer Software, 44, 388-397.

[2]. Raj, S., Ray, K. C., & Shankar, O. (2016). Cardiac arrhythmia beat classification using DOST and PSO tuned SVM. Computer methods and programs in biomedicine, 136, 163-177.

[3]. Vijayarani, S., &Dhayanand, S. (2015). Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816-820.

[4]. Singh, A. S., Irfan, M., &Chowdhury, A. (2018, December). Prediction of liver disease using classification algorithms. In 2018 4th international conference on computing communication and automation (ICCCA) (pp. 1-3). IEEE.

[5]. Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE.

[6]. Senbagavalli, M., &Arasu, G. T. (2016). Opinion mining for cardiovascular disease using decision tree based feature selection. Asian Journal of Research in Social Sciences and Humanities, 6(8), 891-897.

[7]. Valli, M. S., &Arasu, G. T. (2016). An Efficient Feature Selection Technique of Unsupervised Learning Approach for Analyzing Web Opinions.

[8]. Malik, M. R. S., Mohideen, E. R., & Ali, L. (2015, December). Weighted distance grey wolf optimizer for global optimization problems. In 2015 IEEE international conference on computational intelligence and computing research (ICCIC) (pp. 1-6). IEEE.

[9]. Mittal N, Singh U, Sohi BS (2016) Modified grey wolf optimizer for global engineering optimization. Appl Comp Intell Soft Comput. https://doi.org/10.1155/2016/7950348 31.

[10]. Long W, Liang X, Cai S, Jiao J, Zhang W (2017) A modified augmented Lagrangian with improved grey wolf optimization to constrained optimization problems. Neural ComputAppl 28(1):421. https://doi.org/10.1007/s00521-016-2357-x 32.

[11]. Rodrı´guez L, Castillo O, Soria J (2016) Grey wolf optimizer with dynamic adaptation of parameters using fuzzy logic. In: 2016 IEEE congress on evolutionary computation (CEC), pp. 3116–3123. https://doi.org/10.1109/CEC.2016.7744183 33.

[12]. RodrI´guez L, Castillo O, Soria J, Melin P, Valdez F, Gonzalez CI, Martinez GE, Soto J (2017) A fuzzy hierarchical operator in the grey wolf optimizer algorithm. Appl Soft Comput 57:315. https:// doi.org/10.1016/j.asoc.2017.03.048 34.

[13]. Dudani K, Chudasama A (2016) Partial discharge detection in transformer using adaptive grey wolf optimizer based acoustic emission technique. Cogent Eng 3(1):1256083. https://doi.org/10. 1080/23311916.2016.1256083 35.

[14]. Long W, Jiao J, Liang X, Tang M (2018) An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization. EngApplArtifIntell 68:63. https://doi.org/10. 1016/j.engappai.2017.10.024 36.

[15]. Long W, Jiao J, Liang X, Cai S, Xu M (2019) A random opposition-based learning grey wolf optimizer. IEEE Access 7:113810. https://doi.org/10.1109/ACCESS.2019.2934994 37. Sharma S, Salgotra R, Singh U (2017) An enhanced grey wolf optimizer for

numerical optimization. In: 2017 international conference on innovations in information, embedded and communication systems (ICIIECS).

[16]. Sharma, S., Salgotra, R., & Singh, U. (2017, March). An enhanced grey wolf optimizer for numerical optimization. In 2017 international conference on innovations in information, embedded and communication systems (ICIIECS) (pp. 1-6). IEEE.

[17]. Vanderbeck, S., Bockhorst, J., Komorowski, R., Kleiner, D. E., &Gawrieh, S. (2014). Automatic classification of white regions in liver biopsies by supervised machine learning. Human pathology, 45(4), 785-792.

[18]. Dhamodharan, S. (2014, May). Liver disease prediction using bayesian classification. In 4th National Conference on Advanced computing, applications & Technologies (pp. 1-3).

[19]. Ansari, S., Shafi, I., Ansari, A., Ahmad, J., & Shah, S. I. (2011, December). Diagnosis of liver disease induced by hepatitis virus using artificial neural networks. In 2011 IEEE 14th international multitopic conference (pp. 8-12). IEEE.

[20]. Saqlain, S. M., Sher, M., Shah, F. A., Khan, I., Ashraf, M. U., Awais, M., &Ghani, A. (2019). Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. Knowledge and Information Systems, 58, 139-167.

[21]. Latha, C. B. C., &Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203.

[22]. Beunza, J. J., Puertas, E., García-Ovejero, E., Villalba, G., Condes, E., Koleva, G., ...&Landecho, M. F. (2019). Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). Journal of biomedical informatics, 97, 103257.