Wasit Journal of Computer and Mathematics Science
Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM
e-ISSN: 2788-5879 p-ISSN: 2788-5887

Predicting the Optimal Treatment for Diseases Using Whale Optimization Algorithm

Zuhal Adel Madlool[1], Sudad Najim Abed[2]*

[1] Computer science and information technology, Al-Qadisiyah University, Al-Qadisiyah, 58002, IRAQ
[2] Computer science and information technology, Al-Qadisiyah University, Al-Qadisiyah, 58002, IRAQ

*Corresponding Author: Sudad Najim Abed

DOI: https://doi.org/10.31185/wjcms.273
Received 03Augest 2024; Accepted 27 Septemberr 2024; Available online 30 September 2024

ABSTRACT: Recently, there has been a significant research focus on addressing disease prevalence, especially in dealing with the complexities of big data associated with disease data. The challenge is achieving high accuracy due to missing value problems in big data. This study aims to use AI techniques to develop a system that predicates the optimal solutions for disease, regardless of the type of disease, i.e. the system can be applied to any type of disease. The approach involves handling missing values and normalizing disease datasets. The Whale Optimization Algorithm (WOA) will be used to improve predictions for effective disease treatments. We obtained good results for predicting the appropriate treatment for the disease in the proposed research, compared to the results obtained when applying the PSO algorithm before development in state of earlier, where the results obtained in the proposed research had higher accuracy than the results in in state of earlier at high iterations starting from 200 iterations and also had a lower error rate.:

Keywords: Big data, Machine learning, deep learning, Optimization, Prediction

1. INTRODUCTION

Sickness, disease, or illness is a non-affective condition that affects a person or human organs, causing disturbance, weakening of functions, or exhausting the person responsible for his disturbance. The term is sometimes used to denote any characteristic injury, disability, disease, atypical symptoms, deviant control, or typical changes in structure and function, and in other associations, it may be necessary to distinguish between all of these [1].

Many diseases affect humans and the immune system, including chronic and non-chronic diseases. Chronic disease is a disease or condition that is permanent or long-lasting in nature, or develops over time and has slow progresses. It usually lasts three months or longer. The World Health Organization points out those chronic diseases are not transmitted among people [2].

1.1 Big data

Big data is an expression of a large amount of data sets having large, more varied, and complex structures that face storing complexity and analyzing, difficult to visualize for further processes or results [3]. Big data is defined as various sources of data like images, text, and audio. The following factors that make the data big data:

- Size: when the data takes a lot of time to process it.
- Diversity: there are many types of big data structured, unstructured, and half-structured.
- Speed: the frequency speed of data happens, for example, the rapid deployment of tweets is not the same as the rapid of remote sensing of climate change [3].

1.1.1 The importance of big data

Big data is also used to determine the risk factors for diseases and the medical state of patients. Furthermore, many websites and social media provide organizations of healthcare with updated information about the spread of infectious diseases [4]. The following examples explain the use of big data by organizations:

- Big data contributes to the energy industry by identifying suitable locations for drilling and monitoring pipelines and is also used to track electrical grids[5].
- Risk management and real-time analysis using big data and financial service providers.
- For transportation companies: big data is relied upon in the process of managing supply chains for

*Corresponding author: suadad.najim@qu.edu.iq

transportation companies and improving delivery methods[5].

- It has other uses in areas related to the government, such as emergencies and response to them, reducing the occurrence of crimes, as well as contributing to the establishment of smart cities[6].

1.2 Machine learning:

An artificial intelligence technique employed for data analysis utilizes algorithms to simulate human learning, enabling systems to autonomously learn from data and enhance performance without direct human intervention, focusing on practical applications such as prediction and optimization.[7]. It is a scientific discipline centered on two primary objectives: developing computer systems capable of enhancing their performance through practice and experience and understanding the principles governing learning systems through algorithms, data, and theories [8].

- The surge in interest in machine learning can be attributed to its capability to construct models that swiftly and accurately analyze vast and intricate datasets, enabling companies and institutions to mitigate risks and enhance profitability.
- Machine learning algorithms are categorized into three fundamental types: supervised learning, unsupervised learning, and reinforcement learning [9].
- Supervised learning involves predicting future outcomes based on new classified data, achieved by training a model with known inputs and outputs. Data professionals guide algorithms to derive conclusions, akin to teaching a child to recognize animal shapes from a picture book [10].
- On the other hand, unsupervised learning analyzes and clusters unclassified input data without predefined outputs, discovering patterns autonomously. It excels in scenarios lacking historical attributes and labels, such as exploratory data analysis and pattern recognition [10]. Unsupervised techniques like neural networks, k-means clustering, and probabilistic methods like PCA and SVD reduce model dimensions by identifying similarities and differences among data points.
- Reinforcement learning trains models to make decisions by learning from rewards and feedback obtained during interactions with their environment. Unlike supervised learning, reinforcement learning does not rely on sample datasets; instead, algorithms learn through trial and error to maximize rewards over time [11].

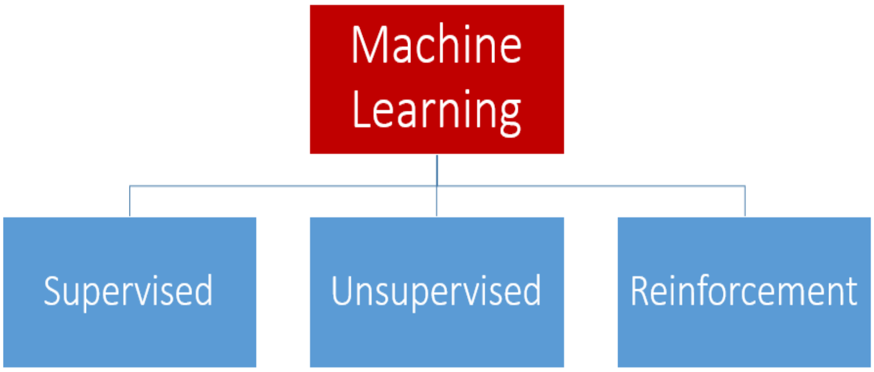

FIGURE 1 Types of Machine Learning

1.2.1 Optimization

Optimization methods are employed across diverse research fields to achieve outcomes that either increase or decrease specific parameters of interest. These methods aim to achieve objectives such as reducing the cost of products or services, increasing profitability, minimizing raw material usage by enhancing quality or improving productivity [12].

Optimization is the process of identifying the best possible solution from a set of feasible solutions. Linear programming and Quadratic programming are examples of optimization types [13].

Optimization algorithms are algorithms that find the optimal solution to a given problem. Many fields use the optimization algorithms such as deep learning and stochastic programming. In this research, one of the optimization algorithms will be developed [12].

1.2.2 Prediction

Predictive algorithms are extensively utilized today for analyzing data structure and evolution using techniques such as data mining and statistics. Various forms of content analytics have been developed based on specific algorithms to achieve desired objectives[14].

Prediction involves using known model parameters to estimate the behavior of a system at specified input states. Both prediction and optimization processes can employ statistical and probabilistic methods. Optimization aims to maximize outcomes with available resources.

Prediction is crucial for forecasting future outcomes and probabilities, as well as anticipating future requirements based on current data and trends[16][15].

1.3 Problem statement:

The problem is to build a system that can predict the optimal solution for any type of disease with high accuracy and less error.

1.4 Contribution:

The contribution of this paper is to use the Whale Optimization algorithm to predict the optimal treatment for different types of diseases, i.e. the system can applied to any type of disease, not to a particular one, and achieves high prediction accuracy. The approach involves efficiently handling missing values and normalizing the dataset to improve the solutions.

1.5 Evaluation strategies:

Disease datasets are used to evaluate the proposed model performance: WOA (Whale Optimization Algorithm). Accuracy, error rate, and MSE are used for unsupervised performance factors. In addition, the accuracy of the proposed model was verified by comparing it with the resulting values of PSO before developed in the paper [16].

1.6 Paper Organization

In section 2, this paper provides an overview of related work that is used for the Whale Optimization Algorithm (WOA). In section 3 the results obtained from the proposed system are presented. A discussion of the results is provided in Section 4. The final section is 5 which includes future research.

2. Related Work

Abed, S. N [16] presented Predicting the Optimal Treatment for Diseases Using the genetic method by developing (PSO) Optimization technique, we will introduce a similar work using the same dataset and evaluation metrics(Error Rate, MSE) to predict optimal treatment for diseases using different algorithm where we obtained highest results in accuracy is(95.00678) than this paper the accuracy of results was(94.07745).

A. Mostafa, Aboul Ella Hassanien, M. Houseni, and H. A. Hefny [17] proposes Liver segmentation in MRI images based on whale optimization algorithm, our work is differ it using another dataset and another evaluation metrics, but is similar to it using the same algorithm to predict the optimal treatment for diseases regardless of the disease.

3. Implementation of Whale Optimization Algorithm

The WOA (Whale Optimization Algorithm) uses intelligence to satisfy its goals by using an effective method for tackling the challenge of simulating the humpback whale's distinctive bubble net feeding strategy. This technique, first introduced by Mirghalili and Lewis in 2016, involves whales trapping prey by creating bubbles in a circular or figure-eight pattern. Humpback whales, which are among the largest mammals on Earth can grow up to 30 meters long and weigh up to 180 tons. There are seven primary species, including Killer, Minke, Sei, Humpback, Right, Finback, and Blue whales, known as apex predators, whales don't sleep; instead, only half of their brain rests at a time, as they need to breathe at the surface.

Whales are very intelligent and have emotional capacity[18].

They have spindle cells in their brains similar to those in humans which are linked to judgment, emotions, and social behavior. Whales possess twice as many of these cells as humans contributing to their notable cognitive abilities. Studies suggest that whales especially killer whales, can think, learn, judge, communicate, and exhibit emotions although to a lesser level of intelligence compared to humans. They are even capable of developing their distinct dialects[19].

Socially whales may live either alone or in groups, though they are often seen in groups. Some species, like killer whales, form lifelong family units. Among the largest baleen whales, humpback whales (Megaptera novaeangliae) are nearly as large as a school bus. The preferred prey of whales includes krill and schools of small fish[20].

The steps adopted to represent the search algorithm are summarized in the following steps:

1-Encircling prey

Humpback whales can recognize the location of prey and encircle them. Since the position of the optimum design in the search space is not known a priori, the WOA algorithm assumes that the current best candidate solution is the target prey or is close to the optimum [21] .After the best search agent is defined, this behavior is represented by the following equations:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_P}(t) - \vec{X}(t)\right| \quad (1)$$

$$\vec{X}(t+1) = \overrightarrow{X_P}(t) - \vec{A}\,.\,\vec{D} \quad (2)$$

Where t indicates the current iteration, $\vec{A}$, $\vec{D}$ are coefficient, $\overrightarrow{X_P}$ vectors is the position vector of the prey, and $\vec{X}$ indicates the position vector of a whale.
The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = \overrightarrow{2a}\,.\overrightarrow{r_1} - \vec{a} \quad (3)$$

$$\vec{C} = 2.\overrightarrow{r_2} \quad (4)$$

Where components of $\vec{a}$ are linearly decreased from 2 to 0 over the course of iterations and r_1,r_2 are random vectors in [0,1].

2-Bubble-net attacking method (exploitation phase)

To mathematical model the bubble net behavior of humpback whales, two methods were designed as follows:

1- Shrinking encircling mechanism: This method is satisfied by reducing the value. of $\vec{a}$. We can notice that the fluctuation range of $\vec{A}$ is also reduced by $\vec{a}$. In other words, $\vec{A}$ is a random value in the interval [-a,a] where a is reduced 0 for iterations. Setting random values for $\vec{A}$ in [-1,1], A new search agent's position is determined somewhere between the original position of the agent and the position of the current best agent.
2- Spiral updating position. The method starts by calculating the distance between the location of the whale at (X,Y) and the location of the prey at (X*, Y*).. The following equation is used to simulate the helix-shaped movement of humpback whales:

$$\vec{X}(t+1) = \overrightarrow{D^1}\, e^{bt} \cos(2\pi t) + \overrightarrow{X^*}(t) \quad (5)$$

Where:

$$\overrightarrow{D^1} = \left|\overrightarrow{X^*}(t) - \vec{X}(t)\right| \quad (6)$$

It also measures the distance of the i-th whale from the prey (the best solution obtained so far), where” b” is a constant that defines the shape of the logarithmic spiral, and "t” is a random number between -1 and 1.We can note that humpback whales use a combination of two techniques to catch their prey, they swim in a shrinking circle around the prey and follow a spiral-shaped path simultaneously. To accurately model this dual behavior we assume there's a 50% chance that the whale will choose either the shrinking circle or the spiral model to update its position during the optimization process. The mathematical representation of this model is as follows:

$$\boldsymbol{X(t+1)} = \begin{cases} \overrightarrow{\boldsymbol{X}} * (\boldsymbol{t}) - \overrightarrow{\boldsymbol{A}}\,\overrightarrow{\boldsymbol{D}} & \boldsymbol{p < 0.5} \\ \overrightarrow{\boldsymbol{D^1}} \boldsymbol{e^{bt}} \,\boldsymbol{cos(2\pi t)} + \overrightarrow{\boldsymbol{X^*}}(\boldsymbol{t}) & \boldsymbol{p < 0.5} \end{cases} \quad \textbf{(7)}$$

where p is a random number between 0 and 1.

3-Search for prey (exploration phase):

The same approach using changes to the $\overrightarrow{\boldsymbol{A}}$ vector can be applied to search for prey. (exploration). Random search is applied by humpback whales according to the position of each other. In the exploration phase of the WOA algorithm, we use the vector$\overrightarrow{\boldsymbol{A}}$ with random values greater than 1 or less than -1 to encourage the search agent to move farther away from a reference whale. Unlike the exploitation phase, where the position is updated based on the best search agent the exploration phase updates the position according to a randomly selected search agent. This approach combined with $|\vec{A}| > 1$,highlights the exploration aspect and supports the algorithm's ability to conduct a global search. The mathematical model is as follows:

$$\vec{D} = |\vec{C}\ \vec{X}_{rand} - \vec{X}| \quad (8)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A}\,\vec{D} \quad (9)$$

where $\vec{X}_{rand}$ is a random position vector, representing a random whale.

Algorithm 1. Whale Optimization algorithm

```
Input data, Number of maxiter and Population etc
Initialize the whales population  Xi (i = 1, 2, ..., n)
Initialize a, A, C, l and p
   //Calculate the fitness of each search agent
X*= the best search agent
while (it < Maxiter)
    for each search agent
        if (p < 0.5)
            if (|A| < 1)
                Update the position of the current search agent by the equation (2)
            else if (|A| ≥ 1)
                Select a random search agent (X_rand)
                Update the position of the current search agent by the equation (9)
            end
        else if (p ≥ 0.5)
            Update the position of the current search agent by the equation (5)
        end
    end
    //Calculate the fitness of each search agent
    //Update X* if there is a better solution
    it=it+1
    Update a, A, C, l and p
end while
return X*
```

After applying the above-proposed algorithm WOA on the database for disease treatments to predict the optimal treatment for the disease, we obtained the results shown in **Table 1**:

Table 1. Result of the Whale Optimization Algorithm

Iteration	Fest fitness value (accuracy)	Error rate	MSE
50	77	8.0042	0.009
100	77.63878	7.9725	0.00885
150	77.53877	7.1077	0.00820
200	82.36466	6. 4185	0.00685
250	82.37454	6. 4185	0.00685
300	82.77466	5.4398	0.00606
350	82.77459	5.4398	0.00606
400	83.57066	5.2898	0.00603
450	83.57067	5.2897	0.00603
500	90. 68889	4.6566	0.00573
550	90. 88798	4.6555	0.00583
600	93.06678	3.6594	0.00305
650	95.00678	2.3465	0.00244
700	95.00678	2.3465	0.00244

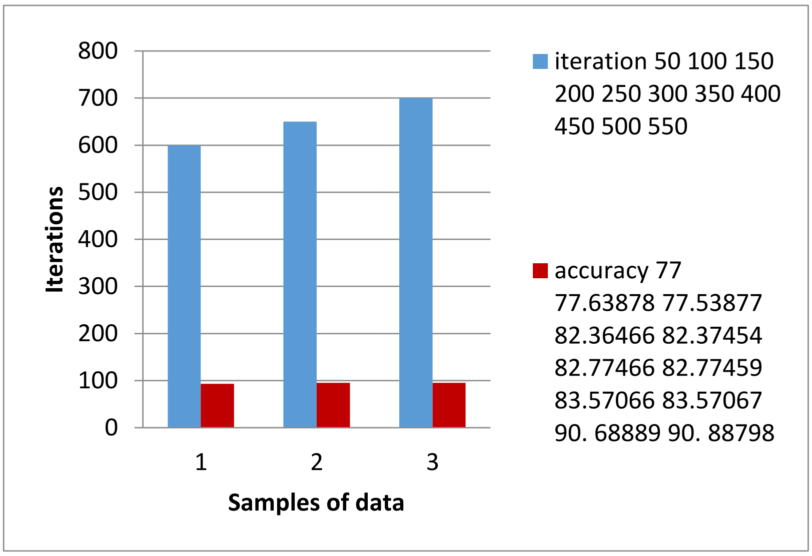

FIGURE 2. Ratio of accuracy

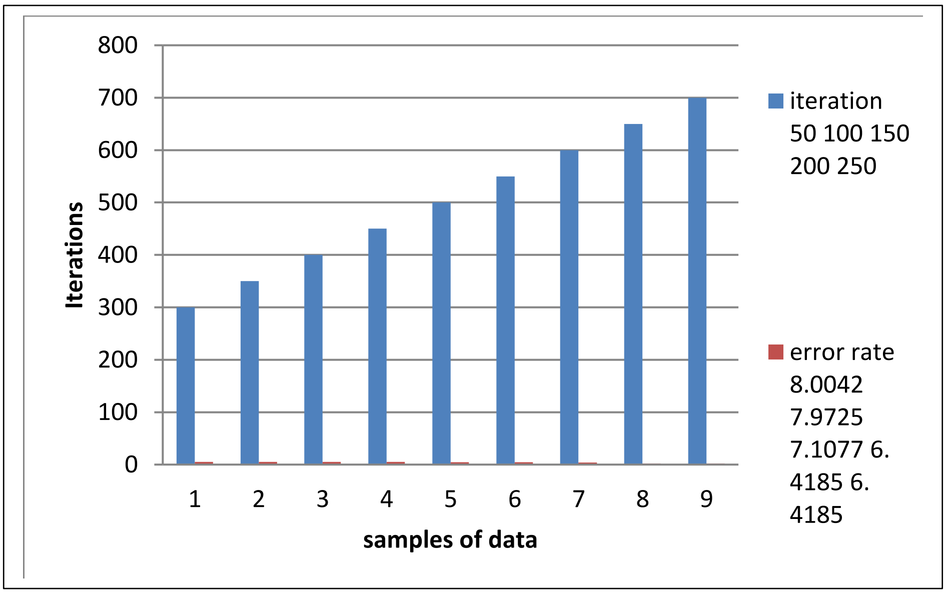

FIGURE 3. Ratio of error rate

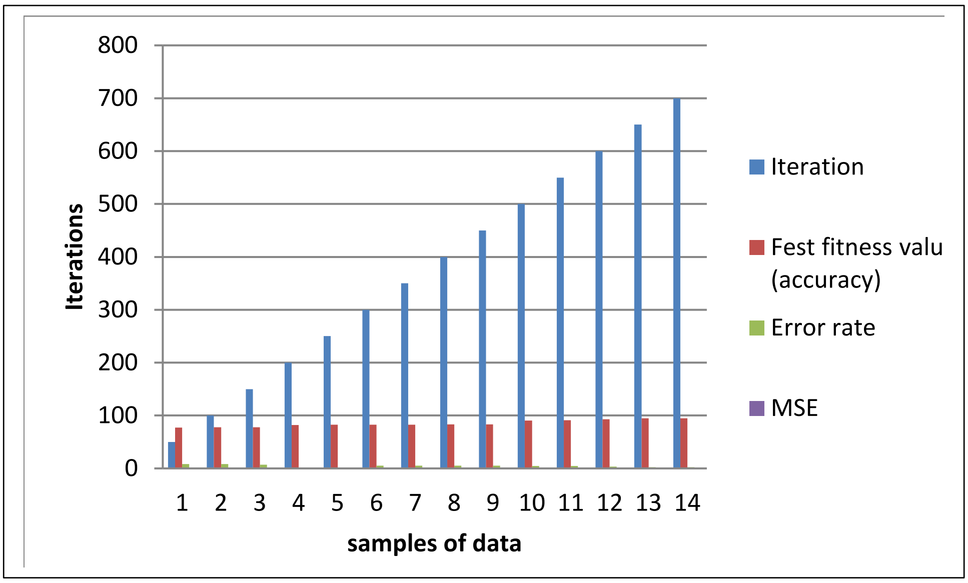

FIGURE 4. Results of WOA algorithm

3.1 Technical proposed system

1- Dataset reading.
2- Handel missing values.
3- Normalization.
4- Apply WOA (Whale Optimization Algorithm) to get Objective functions (Accuracy).
5- Compute error rate and MSE.

The following diagram show the technical steps of the proposed system:

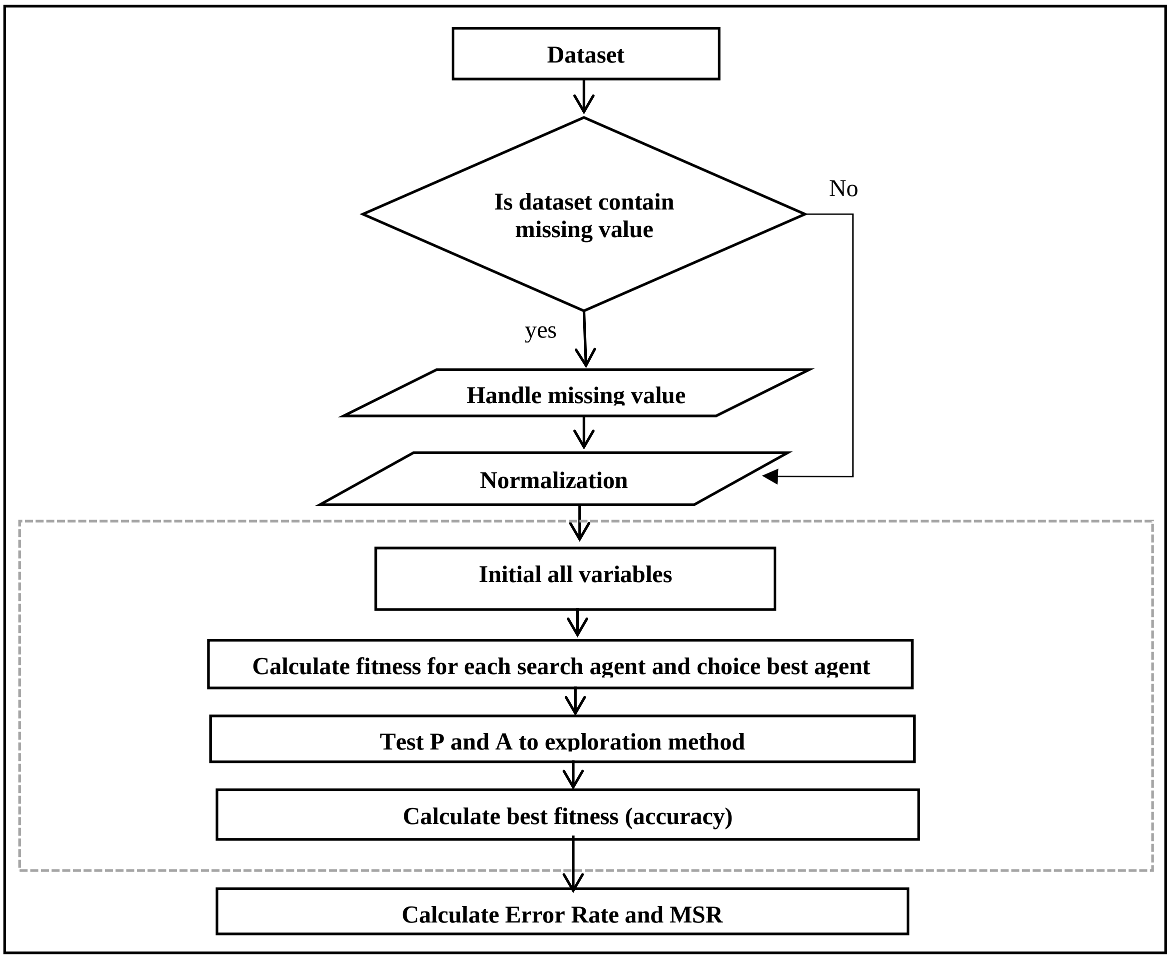

FIGURE 5. The proposed system

3.2 Valuation stage

At this stage, a comparison will be conducted between the results obtained from the proposed system and those achieved using Particle Swarm Optimization (PSO), as detailed in the previous study [16].

In Table 2, we conducted a comparison between the WOA algorithm and PSO as detailed in [16], aiming to determine the superior method. Both tables utilized an identical number of iterations, with errors recorded for every 50 iterations displayed in Table 2.

In the PSO algorithm in Paper [16], the highest accuracy was obtained in iteration 450 and the stop was made at iteration 500, i.e. when there was no change in the accuracy and error values from the previous iteration, while in the proposed research we obtained a higher accuracy at iteration 650 and the stop was made at iteration 700, i.e. when the change in the accuracy and error values from their values in the previous iteration stopped.

Table 2 includes five columns: the first column lists the iterations, while the second and third columns display the error rate and Mean Squared Error (MSE) generated by PSO before development as documented in [16] every 50 iterations. The fourth and fifth columns present the error rate and MSE values obtained from the WOA in the proposed system. Notably, WOA demonstrates superior accuracy starting from iteration 200 compared to PSO before its development in [16].

Table 2. Comparing the error rate between PSO and WOA

Iteration	Results of PSO			Results of WOA		
	Fest fitness value (accuracy)	**Error rate**	**MSE**	**Fest fitness value (accuracy)**	**Error rate**	**MSE**
50	77	8.0043	0.009	77	8.0042	0.009
100	77.63878	7.9826	0.00985	77.63878	7.9725	0.00885
150	77.53877	7.1067	0.00920	77.53877	7.1077	0.00820
200	81.86766	4.8541	0.00585	82.36466	6. 4185	0.00685
250	81.26789	4.0598	0.00506	82.37454	6. 4185	0.00685
300	92.89688	3.6078	0.00473	82.77466	5.4398	0.00606
350	93.26789	3.6098	0.00383	82.77459	5.4398	0.00606
400	93.86678	3.0294	0.00305	83.57066	5.2898	0.00603
450	94.07745	2.8266	0.00277	83.57067	5.2897	0.00603
500	94.07745	2.8266	0.00277	90. 68889	4.6566	0.00573
550	-	-	-	90. 88798	4.6555	0.00583
600	-	-	-	93.06678	3.6594	0.00305
650	-	-	-	95.00678	2.3465	0.00244
700	-	-	-	95.00678	2.3465	0.00244

4. Conclusions

By using the whale algorithm to predict the optimal treatment for diseases, regardless of the type of disease, and after comparing it to the results of the Particle Swarm Optimization algorithm before development described in the state of earlier , and by comparing the percentage of accuracy that we obtained, it became clear that the algorithm gave its benefits in increasing accuracy to predict the appropriate treatment, better than the results of the PSO algorithm before development in the state of earlier, the accuracy is higher at high iterations starting from 200 iterations and also had a lower error rate and MSE.

As future work, the aim is to develop the Whale Optimization Algorithm to achieve more accurate results and less error in the obtained results in this paper.

Funding

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] N. H. Lameire et al., "Harmonizing acute and chronic kidney disease definition and classification: report of a Kidney Disease: Improving Global Outcomes (KDIGO) Consensus Conference," Kidney International, vol. 100, no. 3, pp. 516–526, Sep. 2021, doi: https://doi.org/10.1016/j.kint.2021.06.028.

[2] E. Anderson and J. L. Durstine, "Physical activity, exercise, and Chronic diseases: a Brief Review," Sports Medicine and Health Science, vol. 1, no. 1, pp. 3–10, Sep. 2019, doi: https://doi.org/10.1016/j.smhs.2019.08.006.

[3] S. Sagiroglu and D. Sinanc, "Big data: a Review," 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 42–47, May 2013, doi: https://doi.org/10.1109/cts.2013.6567202.

[4] Ngiam, K. Y., & Khor, W.," Big data and machine learning algorithms for health-care delivery" The Lancet Oncology, 20(5), e262-e273. Article (CrossRef Link)

[5] F. P. S. Surbakti, W. Wang, M. Indulska, and S. Sadiq, "Factors influencing effective use of big data: A research framework," Information & Management, vol. 57, no. 1, p. 103146, Feb. 2019, doi: https://doi.org/10.1016/j.im.2019.02.001.

[6] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," Technological Forecasting and Social Change, vol. 153, p. 119253, Apr. 2018, doi: https://doi.org/10.1016/j.techfore.2018.03.024.

[7] G. Carleo et al., "Machine learning and the physical sciences," Reviews of Modern Physics, vol. 91, no. 4, Dec. 2019, doi: https://doi.org/10.1103/revmodphys.91.045002.

[8] P. Mehta et al., "A high-bias, low-variance introduction to Machine Learning for physicists," Physics Reports, vol. 810, pp. 1–124, May 2019, doi: https://doi.org/10.1016/j.physrep.2019.03.001.

[9] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, Jul. 2020, doi: https://doi.org/10.1126/science.aaa8415.

[10] J. E. van Engelen and H. H. Hoos, "A Survey on semi-supervised Learning," Machine Learning, vol. 109, Nov. 2019, doi: https://doi.org/10.1007/s10994-019-05855-6.

[11] Y. Matsuo et al., "Deep learning, reinforcement learning, and world models," Neural Networks, Apr. 2022, doi: https://doi.org/10.1016/j.neunet.2022.03.037.

[12] S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A Survey of Optimization Methods From a Machine Learning Perspective," IEEE Transactions on Cybernetics, vol. 50, no. 8, pp. 3668–3681, Aug. 2020, doi: https://doi.org/10.1109/tcyb.2019.2950779.

[13] J. R. R. A. Martins and A. Ning, Engineering Design Optimization. Cambridge University Press, 2021. Accessed: Jul. 31, 2024. [Online]. Available: https://books.google.iq/books?hl=en&lr=&id=dBVEEAAAQBAJ&oi=fnd&pg=PR13&dq=%5B13%5D%09Martins.

[14] M. Prosperi et al., "Causal inference and counterfactual prediction in machine learning for actionable healthcare," Nature Machine Intelligence, vol. 2, no. 7, pp. 369–375, Jul. 2020, doi: https://doi.org/10.1038/s42256-020-0197-y.

[15] W. Ben Chaabene, M. Flah, and M. L. Nehdi, "Machine learning prediction of mechanical properties of concrete: Critical review," Construction and Building Materials, vol. 260, p. 119889, Nov. 2020, doi: https://doi.org/10.1016/j.conbuildmat.2020.119889.

[16] S. N. Abed, "Predicting the Optimal Treatment for Diseases Using the Genetic Method by Develop (PSO) Optimization Technique," Journal of Al-Qadisiyah for Computer Science and Mathematics, vol. 16, no. 2, Jul. 2024, doi: https://doi.org/10.29304/jqcsm.2024.16.21545.

[17] A. Mostafa, Aboul Ella Hassanien, M. Houseni, and H. A. Hefny, "Liver segmentation in MRI images based on whale optimization algorithm," Multimedia Tools and Applications, vol. 76, no. 23, pp. 24931–24954, Apr. 2017, doi: https://doi.org/10.1007/s11042-017-4638-5.

[18] S. Mahmood, N. Z. Bawany, and M. R. Tanweer, "A comprehensive survey of whale optimization algorithm: modifications and classification," Indonesian Journal of Electrical Engineering and Computer Science, vol. 29, no. 2, p. 899, Feb. 2023, doi: https://doi.org/10.11591/ijeecs.v29.i2.pp899-910.

[19] M. H. Nadimi-Shahraki, H. Zamani, Zahra Asghari Varzaneh, and Seyedali Mirjalili, “A Systematic Review of the Whale Optimization Algorithm: Theoretical Foundation, Improvements, and Hybridizations,” Archives of Computational Methods in Engineering, May 2023, doi: https://doi.org/10.1007/s11831-023-09928-7

[20] P. L. Tyack, “Social Organization of Baleen Whales,” pp. 147–175, Jan. 2022, doi: https://doi.org/10.1007/978-3-030-98449-6_7

[21] J.-O. Meynecke et al., “The Role of Environmental Drivers in Humpback Whale Distribution, Movement and Behavior: A Review,” Frontiers in Marine Science, vol. 8, Nov. 2021, doi: https://doi.org/10.3389/fmars.2021.720774.