



Available online at [www.qu.edu.iq/journalcm](http://www.qu.edu.iq/journalcm)

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



# Predicting the Optimal Treatment for Diseases Using the Genetic Method by Develop (PSO) Optimization Technique

Sudad Najim Abed

Computer science and information technology, Al-Qadisiyah University, Al-Qadisiyah, Iraq. [e-mail : [suadad.najim@qu.edu.iq](mailto:suadad.najim@qu.edu.iq)]

## ARTICLE INFO

### Article history:

Received: 26 /05/2024

Revised form: 23 /06/2024

Accepted : 27 /06/2024

Available online: 30 /06/2024

### Keywords:

Big data,  
Machine learning,  
deep learning,  
Optimization,  
Prediction,  
Genetic algorithms.

## ABSTRACT

Recently, most researchers have been interested in finding appropriate solutions to the spread of diseases, and the data on diseases considered big data, so dealing with big data is difficult in terms of obtaining highly accurate results because of the missing values. To enhance the deals with this type of data, this paper will use artificial intelligence techniques to design a system to predict the optimal treatment for diseases, regardless of disease type with normalization and processing missing data for dataset of diseases . The Particle Swarm Optimization (PSO) algorithm used to find the best solutions since genetic algorithms have characteristics that distinguish them from other algorithms such as mutation and crossover. The 1X method of crossover has been used to develop the PSO algorithm to give the best prediction for the appropriate treatment for these diseases by using the mathematical equations that are explained in 1.3.1. To prove the enhancement of the PSO algorithm after development, the results obtained from the PSO algorithm were compared before and after development by measuring the Best Fitness Value (accuracy), Error rate, and MSE measurement, the results displayed improvement after development of the algorithm by increased values of accuracy reached to 94% and decreased of Error rate and MSE.

MSC..

<https://doi.org/10.29304/jqcm.2024.16.21545>

## 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, as there are chronic diseases 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].

Chronic diseases cause 60% of deaths in the world. In poor countries, chronic disease deaths reach 80%. Chronic disease is responsible for the death of young people who are less than 70 years old. Around the world, equal effects of chronic diseases on men and women.

\*Corresponding author

Email addresses:

Communicated by 'sub editor'

### 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 :

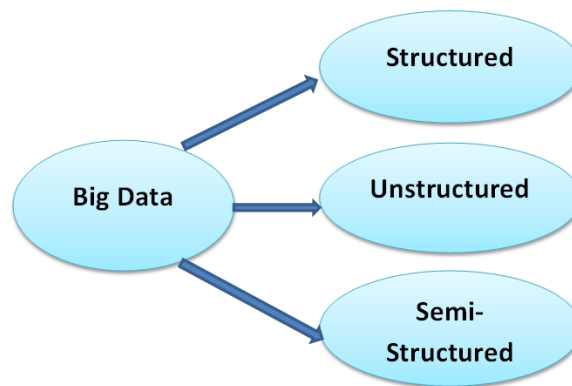
- Size: when the data takes a lot of time to process it
- Diversity: there are many types of big data structured, unstructured, and half-structured.
- Speed: the frequency speed of data happens, for example, the rapid deployment of tweets is not the same as the rapid of remote sensing of climate change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

#### 1.1.1 Big data Types

Big data is a collection of structured, semi-structured, and unstructured data collected by organizations that can be mined for data and used in machine learning, modeling, and other analytical applications. [3]. Big data also contains many data types, including the following:

- Structured data, like relational databases;
- Unstructured data, like web text, and legal documents;
- Semi-structured data, emails, zipped files.

Systems for using and processing big data have become a dominant part of information management organizations with the support of big data analytics tools. [4].



**Fig. 1.** Types of Big Data

#### 1.1.2 Why big data is important?

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
- 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 transportation companies and improving delivery methods.
- 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.

#### 1.1.3 Big Data Challenges

Regarding processing power issues, the architecture design of big data is a big challenge for users. Big data systems must be tailored to meet an organization's specific requirements, Information technology and data management teams are required for DIY to collect specified technology sets and tools. Also, new abilities compared to those typically owned by

database managers and developers who focus on relational software are required by big data systems management and deployment. [5].

Both of these issues can be mitigated by using a managed cloud service, but IT managers need to closely monitor cloud usage to guarantee costs don't get out of control. Migrating on-premises datasets and processing workloads to the cloud is a complex operation.

There are other challenges in a large data management system, such as providing scientists and analysts with access to data, especially in a distributed environment with disparate websites and data warehouses. To give analysts access to relevant data, data management and analysis teams create data catalogs that include metadata management and online work. Integrating the big data process is also often difficult, especially when things happen differently and fast. [6].

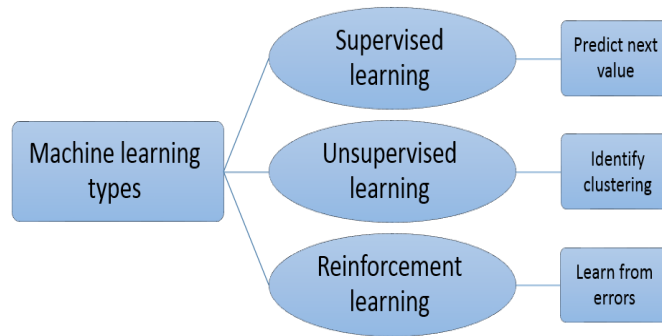
### ***1.2 Machine learning:***

An artificial intelligence technique that is used for the analysis of data using algorithms by simulating the way learning humans to make systems able to learn automatically from data and improve the experience without human interference as possible and achieve practical goals and applications, especially prediction and optimization [7]. Machine learning has increasingly grown in many aspects of new technologies [8]. Machine learning is a discipline of science that focuses on achieving two main goals: how to build computer systems that can improve their work through practice and experience, and what laws govern learning systems from calculations, information, and theories [8].

Many factors contributed to the increase in interest in machine learning, which led to the possibility of obtaining models capable of analyzing a large amount of data that may be complex very quickly and with high accuracy, and this leads companies and institutions to avoid risks and obtain more profits.

Algorithms of Machine learning are based on three basic categories, which are supervised learning, unsupervised learning, and reinforcement learning [9].

- Supervised learning: a technique in which future outcomes are predicted in response to new classified data by training a model with known inputs and outputs. The most commonly used technique is supervised learning, where data professionals guide, teach, and train algorithms on the conclusion that the algorithm should provide, just like teaching a child to identify animal shapes by memorizing them from a picture book[10].
- Unsupervised learning: This technique uses unclassified input data, analyzes and aggregates it, and is not provided with the specified outputs as in supervised learning. It trains and learns by itself to discover evidence of the required data. This type uses a more independent approach, in which complex processes and patterns are identified by the computer without human supervision. This technique is used if the data does not have historical attributes and labels, unlike supervised learning.
- Unsupervised machine learning algorithms are an ideal solution for exploratory data analysis, cross-selling strategies, and image and pattern recognition because they can compare information and detect similarities and differences among them. It also reduces the dimensions of the model making it able to reduce feature numbers. Supervised learning is similar to the process of a child learning different objects through their colors and patterns instead of relying on memorizing names, as this will prompt the child to search for similarities between images and classify them. Examples of unsupervised machine learning algorithms include neural networks, k-means clustering, and probabilistic clustering methods. There are two more common approaches to this technique: PCA and SVD.
- Reinforcement learning: In this type of machine learning, the model is trained on how to make a set of decisions based on the rewards and responses obtained. The machine learns how to achieve goals in complex and uncertain situations as it gets rewarded whenever a goal is achieved during learning. This type of learning is similar to supervised learning but the difference is its algorithm does not have sample data to train in other words it does not provide answers, results, or outputs. So the steps will be performing the task by the reinforcement agent. In other words, the algorithm of this type will detect the procedures that achieve the greatest reward by trial and error, and the algorithm will learn from the experiments it performs if it does not have a training data set. The agent must choose the procedures that increase the expected reward during a certain period [11].



**Fig. 2.** Types of Machine Learning

### 1.2.1 Optimization

Optimization methods are used in many research fields to obtain results that increase or decrease certain research parameters, such as reducing the cost of a good product or service, increasing profits, reducing raw materials by improving quality or increasing productivity [12]. The optimization definition is to find the best possible solutions among many feasible solutions. Linear programming and Quadratic programming are 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 widely used today as they help determine the structure and evolution of given data using data mining and statistical techniques. Different types of content analytics have been developed depending on the algorithm to accomplish the desired tasks.[14]

Prediction implies using the known model parameters to estimate system output behavior at some specified input state. Both processes can use statistical and probabilistic methods. Optimization is obtaining the best out of available resources. To forecast the future and the probability of an outcome and future requirements, prediction be used [15].

## 1.3 Genetic algorithm

We know that an algorithm is a set of sequential instructions aimed at solving a problem or achieving a specific goal. A genetic algorithm is one of the random search methods used in computing, to provide the right solutions or the closest to the optimal solution called search space, by using a comparison among a set of steps and calculating the “distance” between the solutions and then choosing appropriate solutions and depending on them to find other solutions.

During the application of the genetic algorithm, describing the problem is done by representing the chromosomes that represent the solutions in one of the coding methods. As a next step, a set of mathematical operations derived from the biological processes CrossOver, Selection, and Mutation are applied to finally obtain a set of chromosomes that represent the last generation. By choosing the best chromosome, the optimal solution that we want to search for in the issue at hand is reached. This search algorithm starts from a set of random solutions and ends by selecting one from the set of “sons” of solutions.

### 1.3.1 Elements of Genetic Algorithms

Although genetic algorithms differ according to the branches of evolutionary computing, they share at least the following elements:

- (1) Chromosomes of Populations: This represents the search group or Space Search, which is a group of solutions to the issue [16].
- (2) Selection: It consists of choosing the appropriate chromosomes as "parents" to perform the process of intermarriage of them. The selection process for these chromosomes does not happen randomly, but it depends on the capability of the chromosome (Fitness) [17].

- (3) Fitness: It is a coefficient that gives each chromosome a certain value that indicates the efficiency of the chromosome (its approach to the solution, and accordingly the process of selecting the chromosomes takes place [17].
- (4) Crossover: After selecting the appropriate chromosomes from the first generation, depending on the mother's chromosomes, new chromosomes (new offspring) are formed through crossing over. There are three types of crossover, 2X, and UX [16].
- (5) Mutation Random: The changes in chromosomal formulas, which are called mutations, are made after the formation of new offspring, and this helps to faster reach of solution [17].

#### **1.4 Problem statement:**

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

#### **1.5 Contribution:**

The main contribution of this paper is to develop an approach to predict the optimal treatment for diseases, regardless of disease type with high prediction accuracies, by developing the Particle Swarm Optimization (PSO) algorithm using 1X genetic method of crossover, handling missing values and normalize the dataset to find the best solutions.

#### **1.6 Evaluation strategies:**

To estimate the value of optimization of the performance of the proposed system, recent work has been developed using genetic algorithms. Disease datasets are used in order to evaluate the proposed model performance: DPSO (Developed Particle Swarm Optimization). Accuracy, error rate, MSE: Three measures of unsupervised performance factors are used. In addition, the accuracy of the proposed model was verified by comparing it with the proposed system before development.

#### **1.7 Paper Organization**

Section 2 of the paper explains the related work, implementing bird swarm optimization (PSO). The results of the proposed system before and after development are presented in Section 3; a discussion of the results in Section 4. Conclusions and Future Work in Section 5.

## **2. Related Work**

Samaheral-janabi, Mustafa Mohammed, and Ali Al-Sultan,2020[18] designed a smart system based on the PSO algorithm to predict the proportions of gases that cause air pollution, we will introduce a similar work using the same algorithm, but that algorithm is developed by using one of the methods of the genetic algorithm to predict the appropriate treatment for diseases. Our enhancement was increasing the accuracy and decreasing the Error rate and MSE.

In [19] An Effective Wireless Sensor Network Routing Protocol Based on Particle Swarm Optimization Algorithm. The PSO algorithm "Particle Swarm Optimization" has been used to propose an optimization approach in this paper in order to develop multipath protocol called "Particle Swarm Optimization Routing Protocol (MPSORP)". Our work is differing it by developing the PSO algorithm using one of the methods of the genetic algorithm to predict the appropriate treatment for diseases.

In [20], a hybrid genetic algorithm (GA) and particle swarm optimization (PSO) optimized approach based on random forest (RF), called GAPSO-RF, is developed and used to select the optimal features that can increase the accuracy of heart-disease prediction. The proposed system, as a first step excutes multivariate statistical analysis to select the most significant features used in the initial population. As a second step, a discriminate mutation strategy implemented in GA. GAPSO-RF combines a modified GA for global search and a PSO for local search. Moreover, PSO achieved the concept of rehabbing individuals that had been refused in the selection process.

The performance of the proposed system measured by evaluation metrics, namely, accuracy, specificity, sensitivity, and area under the receiver operating characteristic (ROC) curve by using two datasets from the University of California, namely, Cleveland and Statlog. The experimental results recorded accuracies of 95.6% and 91.4% on the Cleveland and Statlog datasets, respectively. Our proposed system is differing by it designed for predicating appropriate treatment for any type of disease not for specific type of disease, and the common evaluation metric used is the accuracy that recorded to 94% in our proposed system.

### 3. Implementation of Particle Swarm Optimization(PSO)

"Swarm" is a set of elements that the used algorithm depends on, distributed in a random way in a limited space to search for the optimal solution in this specific area. First, we will start by enumerating and mentioning some introductory points about PSO:

Population is a grouping, and this grouping depends mainly on the techniques of examples based on statistics. The assembly, in turn, is a set of elements, which represent - within any system that uses the algorithm - a set of solutions, and we will call each element a particle.

Dr. Eberhart and Dr. Kennedy developed this model in 1995, and this model inspired by herd behavior of birds or groups of fish while moving from one place to another.

There are many similarities between PSO and other evolutionary computation techniques. For example, "Genetic Algorithm GA", since the goal of this algorithm is to obtain the optimal and best solution and result - by simulating the behavior of birds in the search for better food. Any system that relies on this algorithm will initially be formed from a random pool of random solutions, and within this pool, the optimal solution will be searched through updating generations.

In effect, the searching operation for the optimal solution by tracking done by the PSO algorithm, and following the current best particles - similar to the behavior of bees and ants, for example.

Comparing the PSO algorithm with GA genetic algorithms, the PSO algorithm is easy to implement. Anyway, the PSO algorithm differs from genetic algorithms, because PSO does not have evolutionary tools such as crossover and mutation, and for this reason, we will develop the PSO algorithm using one of the crossover methods.

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

- Step one: Enter data that is divided into training data and test data.
- Step two: Determine the matrix that is used to extract specifications.
- Step three: start implementing the particle swarm optimization algorithm on the training data.
- Step Four: Set the velocities and locations with initial values, where the velocities are zero at the beginning.
- Step Five: Calculate inertia:

$$\theta = \theta_{max} - \frac{\theta_{max} - \theta_{min}}{i_{max}} i \quad (1)$$

- Step six: Calculate Update velocity and Update position in each particle according to the following equation:
  - Update velocity:

$$u_i = \theta u_{i-1} + c_1 r_1 [x^* - x_{i-1}] + c_2 r_2 [g^* - x_{i-1}] \quad (2)$$

- Update position:

$$x_i = x_{i-1} + u_i \quad (3)$$

- Step seven: find the new position according to f by (check/ Update: X\*, g\*).
- Step eight: Check for convergence.

a particle swarm has been developed to find the best solution for prediction by achieving one of the crossover methods called the 1X method Moderation solution has been generated as follows:

$$Y = w_1 x_1 + w_2 x_2 + \dots + w_i x_i \quad (4)$$

#### Algorithm 1. Standard particle Swarm Optimization

Inputs: N, x1 ,x2 , c1, c2 , imax, f

Output: A swarm S of size N(N position vectors)

Initialize S, randomly generate the position X of each particle w.r.t. The bounds x1, xu of the objective function;

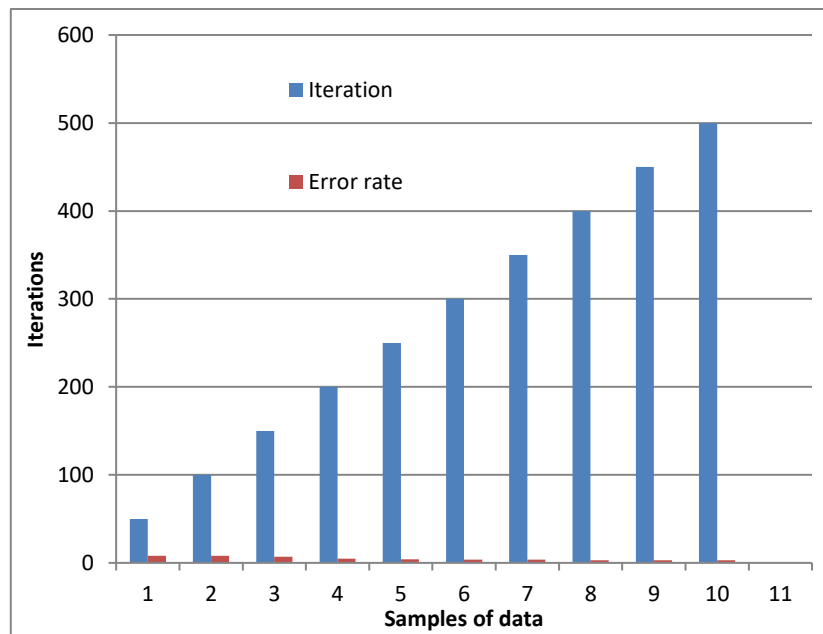
Initialize all velocities u to zero;

Initialize best positions  $x^*$  ( and respective values) for individual particles and find  $g^*$ ;  
 Choose randomly two values in  $[0,1]$  for  $r_1$  and  $r_2$ ;  
 Iteration  $i=0$ ;  
 Initialize  $\theta_{min}$ ,  $\theta_{max}$  ;  
 1: while  $i < i_{max}$  do  
 2: Calculate inertia:  $\theta = \theta_{max} - (\theta_{max} - \theta_{min}) / i_{max} \cdot i$ ;  
 3: For each particle in S values for iteration  $i$  are :  
 4: Update velocity:  $u_i = \theta \cdot u_{(i-1)} + c_1 \cdot r_1 [X^* - X_{(i-1)}] + c_2 \cdot r_2 [g^* - X_{(i-1)}]$   
 5: Update position :  $X_i = X_{(i-1)} + u_i$ ;  
 6: Compute the value of the new position according to  $f$ :  
 7: check/ Update :  $X^*$  ,  $g^*$   
 8: End for  
 9: (Optional) convergence checking;  
 10: Upgrade iteration:  $i=i+1$ ;  
 11: End while  
 12: return S;

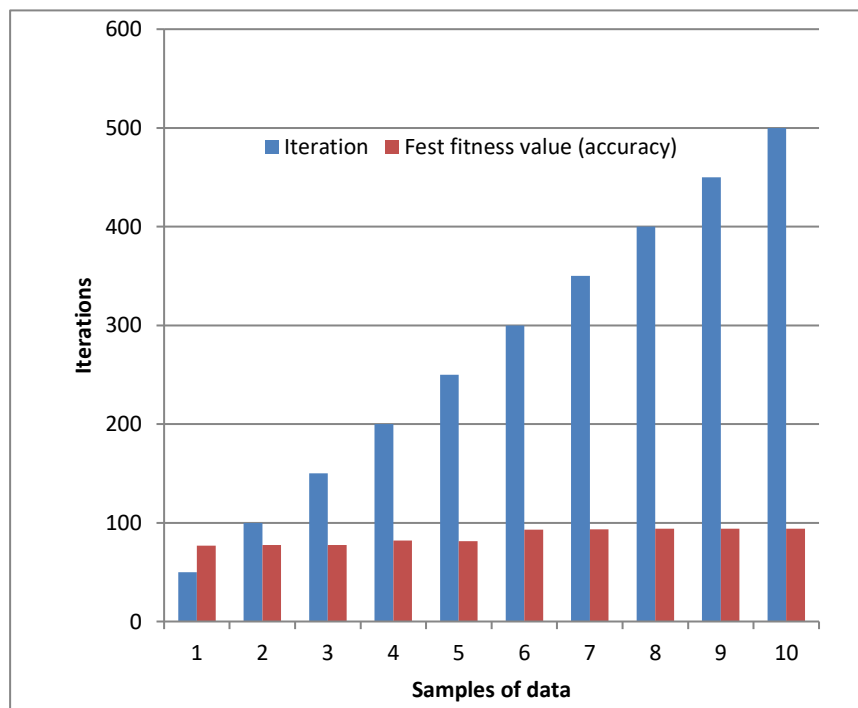
After applying the above-proposed algorithm PSO on the database for disease treatments to predict the optimal treatment for the disease, we obtained the results shown in **Table 1**:

**Table 1.** Result of the Standard Particle Swarm Optimization Algorithm before Developing:

Iterati on	Fest fitness value (accuracy)	Error rate	MSE
50	77	8.0043	0.009
100	77.63878	7.9826	0.00985
150	77.53877	7.1067	0.00920
200	81.86766	4.8541	0.00585
250	81.26789	4.0598	0.00506
300	92.89688	3.6078	0.00473
350	93.26789	3.6098	0.00383
400	93.86678	3.0294	0.00305
450	94.07745	2.8266	0.00277
500	94.07745	2.8266	0.00277



**Fig. 3.** Ratio of error rate



**Fig. 4.** Ratio of Accuracy



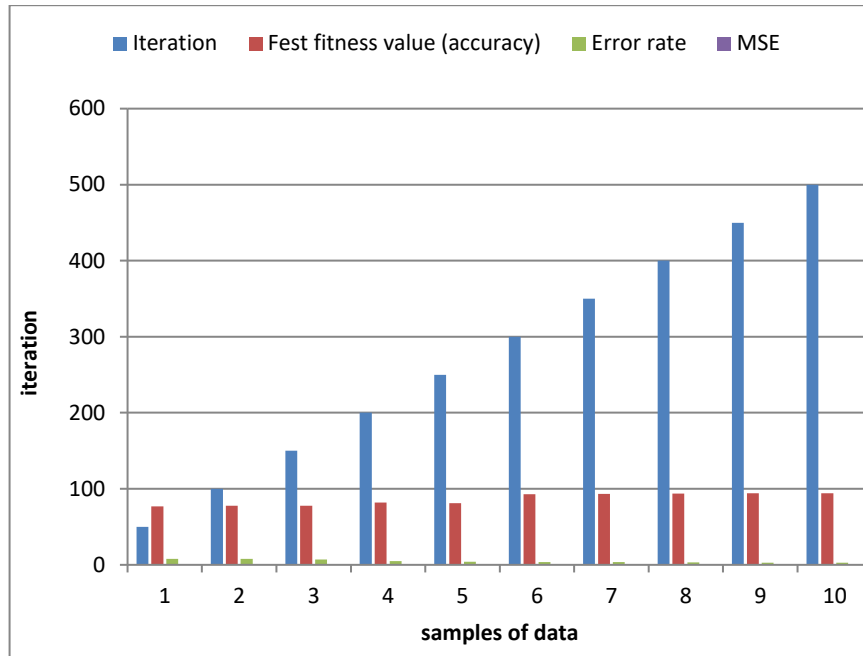


Fig. 5. Results of PSO (before development)

### Algorithms 2: DPSO (Developed Particle Swarm Optimization )

**Inputs:**  $N, x_1, x_2, c_1, c_2, i_{max}, f$

**Output:** A swarm  $S$  of size  $N$  ( $N$  position vectors)

Initialize  $S$ , randomly generate the position  $X$  of each particle w.r.t. The bounds  $x_l, x_u$  of the objective function;

Initialize all velocities  $u$  to zero;

Initialize best positions  $x^*$  ( and respective values) for individual particles and find  $g^*$ ;

Choose randomly two values in  $[0,1]$  for  $r_1$  and  $r_2$ ;

Iteration  $i=0$ ;

Initialize  $\theta_{min}, \theta_{max}$  ;

1: For each sample  $I //i=1..n$

2: If  $i$  have a missing value

3: For each feature  $j //J=1..m$

4: //handle missing value

5: if  $a[i,J]=$  “ “

$a[i,J]=\text{mean}(1/n * \sum XJ)$

6: else

7: return  $a[I,J]$

8: endif

9: endfor

10: else

// Normalization

11:  $a[i,J]=(a[i,J]-\min[i])/(\max[i]-\min[i])$

12: endif

13: endfor

14: while  $i < i_{max}$  do

15: Calculate inertia:  $\theta = \theta_{max} - \frac{\theta_{max} - \theta_{min}}{i_{max}} i$

16: For each particle in  $S$  values for iteration  $i$  are :

17: Update velocity:  $u_i = \theta u_{i-1} + c_1 r_1 [X^* - X_{i-1}] + c_2 r_2 [g^* - X_{i-1}]$

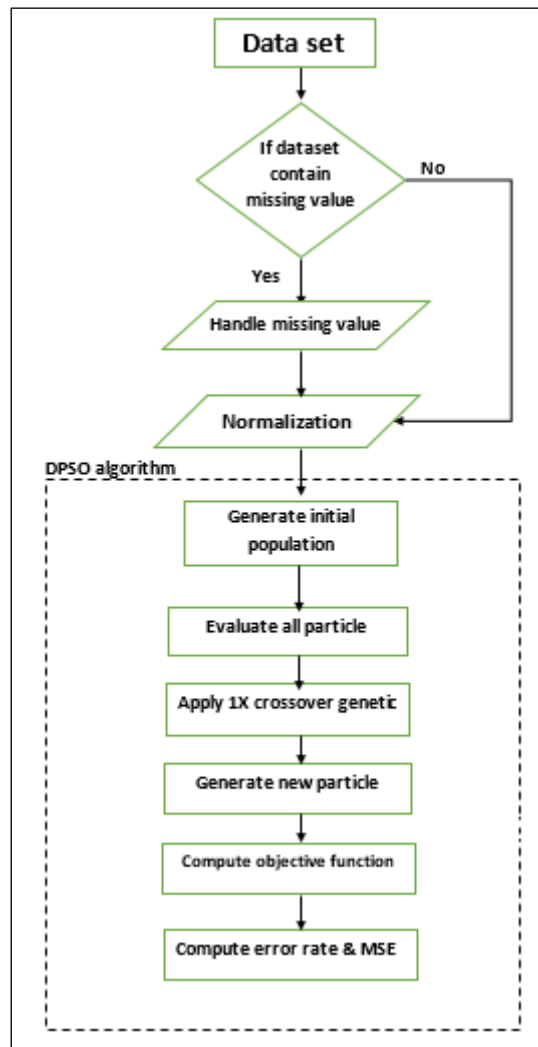
18: Update position :  $X_i = X_{i-1} + u_i$ ;

19: apply Crossover developing  
 20: For each particle in S values for iteration i are:  $Y = w_1x_1 + w_2x_2 + \dots + w_i x_i$   
 21: Compute the value of the new position according to f after develop:  
 22: check/ Update :  $X^*$ ,  $g^*$   
 23: End for  
 24: (Optional) Check for convergence;  
 25: Upgrade iteration:  $i=i+1$ ;  
 26: End while  
 27: return S;

### 3.1 Technical proposed system

- 1- Dataset reading.
- 2- Missing values processing.
- 3- Normalization.
- 4- Apply DPSO(Developed PSO algorithm)
- 5- Compute error rate and MSE.

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

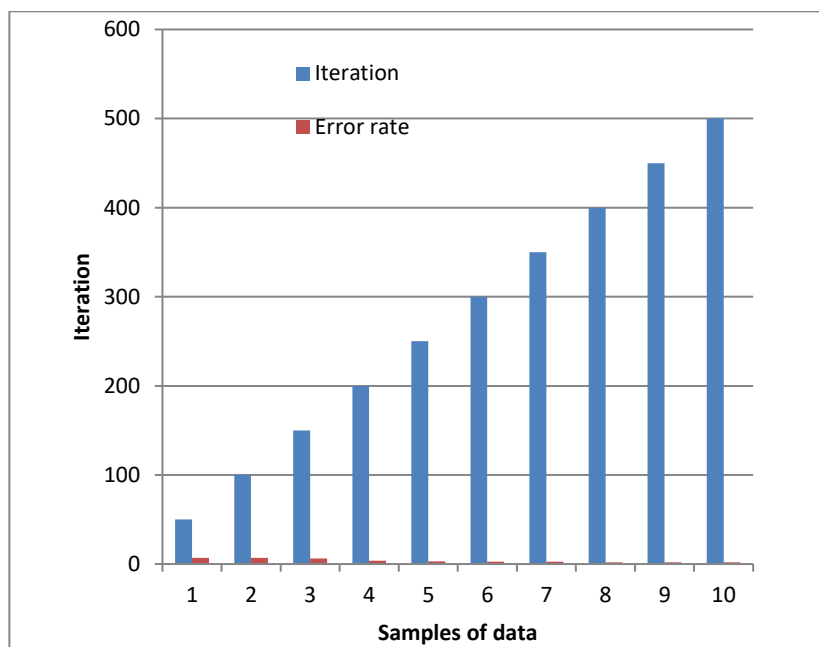


**Fig.6.** The proposed system

After applying the above developed proposed algorithm DPSO on the database, we obtained the results shown in **Table 2**:

**Table 2.** New Result after Applying Crossover Develop

Iteration	Fest fitness value (Accuracy)	Error rate	MSE
50	88	7.0044	0.009
100	89.73838	6.8918	0.00876
150	89.35798	6.1078	0.00810
200	92.97893	3.7462	0.00474
250	92.27889	3.0586	0.00405
300	93.89599	2.5066	0.00362
350	94.26789	2.6068	0.00271
400	94.95678	2.0294	0.00203
450	95.08957	1.9155	0.00187
500	95.08957	1.9155	0.00187

**Fig. 7.** Error Rate

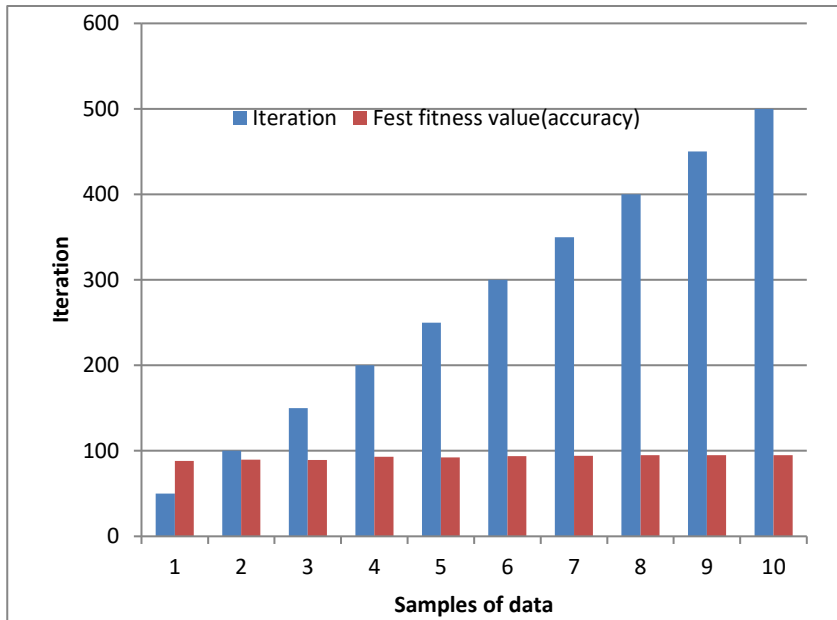


Fig. 8. Ratio of Accuracy

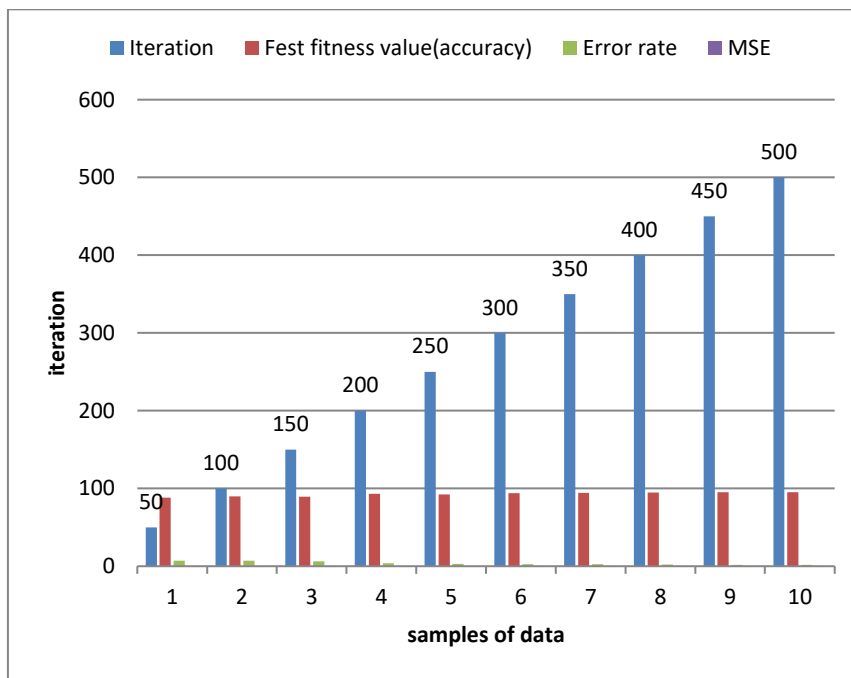


Fig. 9. Results of DPSO (after development)

### 3.2 Valuation stage

At this stage, a comparison will be made for the error to both PSO & DPSO, and this is explained in **Table 1** and **Table 2**.

In **Table 3**, we made a comparison between the algorithm before and after development to determine the best one of them. The same number of iterations is used in both tables and the error for every 50 iterations is shown in **Table 3**, which consists of five columns: the first column contains the iterations, the second and third columns represent the error rate, and the MSE resulted from the PSO before develop in every 50 iterations, the fourth and fifth columns contain the error rate and MSE values resulted from DPSO after develop, as a result, DPSO gave a better result than PSO.

**Table 3:** Comparing the error rate between PSO and DPSO

Iteration	Results of PSO		Results of DPSO	
	Error rate	MSE	Error rate	MSE
50	8.0043	0.009	7.0044	0.009
100	7.9826	0.0098	6.8918	0.0087
150	7.1067	0.0092	6.1078	0.0081
200	4.8541	0.0058	3.7462	0.0047
250	4.0598	0.0050	3.0586	0.0040
300	3.6078	0.0047	2.5066	0.0036
350	3.6098	0.0038	2.6068	0.0027
400	3.0294	0.0030	2.0294	0.0020
450	2.8266	0.0027	1.9155	0.0018
500	2.8266	0.0027	1.9155	0.0018

#### 4. Conclusion

It is evident from the use of the particle swarm optimization al. that developed using the crossover in the genetic algorithm to predict the optimal treatment for disease After comparing the results of the Particle Swarm Optimization algorithm shown in **Table 1** and **Table 2** before and after development, and by comparing the percentage of accuracy that we obtained, it is clear that the algorithm gave its benefits in increasing accuracy to predict the appropriate treatment. We found that the results after development are better, as the accuracy has become higher and the error rate and MSE are less.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest.

#### References

- 1- Disease Definition & Meaning - Merriam-Webster. Accessed 15/11/2023.DOI: Disease Definition & Meaning - Merriam-Webster
- 2- Anderson, E., & Durstine, J. L. (2019). Physical activity, exercise, and chronic diseases: A brief review. *Sports Medicine and Health Science*, 1(1), 3-10. Article (CrossRef Link)
- 3- Sagioglu, S. and Sinanc, D., 2013, May. Big data: A review. In 2013 international conference on collaboration technologies and systems (CTS) (pp. 42-47). IEEE. DOI: Article (CrossRef Link)
- 4- Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262-e273. Article (CrossRef Link)
- 5- Guo, J., Yu, H., Xing, S., & Huan, T. (2022). Addressing big data challenges in mass spectrometry-based metabolomics. *Chemical Communications*, 58(72), 9979-9990. Article (CrossRef Link)
- 6- Khan, M. I., Khan, S., Khan, U., & Haleem, A. (2023). Modeling the Big Data challenges in context of smart cities--an integrated fuzzy ISM-DEMATEL approach. *International Journal of Building Pathology and Adaptation*, 41(2), 422-453. Article (CrossRef Link)
- 7- Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L. and Zdeborová, L., 2019. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4), p.045002. Article (CrossRef Link)
- 8- Mehta, P., Bukov, M., Wang, C.H., Day, A.G., Richardson, C., Fisher, C.K. and Schwab, D.J., 2019. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports* 810, pp.1-124. Article (CrossRef Link)
- 9- Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), pp.255-260. :Article (CrossRef Link).
- 10- Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine learning*, 109(2), 373-440. Article (CrossRef Link)
- 11- Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., ... & Morimoto, J. (2022). Deep learning, reinforcement learning, and world models. *Neural Networks*, 152, 267-275. Article (CrossRef Link)

- 12- Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2019). A survey of optimization methods from a machine learning perspective. *IEEE transactions on cybernetics*, 50(8), 3668-3681. Article (CrossRef Link)
- 13- Martins, J. R., & Ning, A. (2021). *Engineering design optimization*. Cambridge University Press.
- 14- Prospero, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., ... & Bian, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7), 369-375. Article (CrossRef Link)
- 15- Chaabene, W. B., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, 260, 119889. Article (CrossRef Link)
- 16- Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia tools and applications*, 80, 8091-8126. Article (CrossRef Link)
- 17- Lambora, A., Gupta, K., & Chopra, K. (2019, February). Genetic algorithm-A literature review. In *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 380-384). IEEE. Article (CrossRef Link)
- 18- Samaher al-janabi, Mustafa Mohammed, Ali Al-Sultan(2020) Anew method for prediction of air pollution based on intelligent computation. *Soft Computing* 24(1),661-680. Article (CrossRef Link)
- 19- Ghawry, M. Z., Amran, G. A., AlSalman, H., Ghaleb, E., Khan, J., Al-Bakhrani, A. A., ... & Ullah, S. S. (2022). An Effective wireless sensor network routing protocol based on particle swarm optimization algorithm. *Wireless Communications and Mobile Computing*, 2022. Article (CrossRef Link)
- 20- El-Shafiey, M.G., Hagag, A., El-Dahshan, ES.A. et al. A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. *Multimed Tools Appl* 81, 18155-18179 (2022). Article (CrossRef Link)