A hybrid model to optimize medical data using the PSO algorithm

Safaa J. Mohammed
Diwaniyah Directorate of Education
saffa@gmail.com
+9647718884035

Abstract:
The problems of high-dimensionality of medical data are considered the most important thing in any dataset analysis,4cially that which specializes in high-dimensional medical data, where you cannot determine which feature falls among the thousands of features that are located on each chromosome and in each of them are thousands of genes. Reducing the dimensions of high-dimensional medical data is an important matter because it has a great impact on the early diagnosis of many diseases, including all kinds of cancer, and most importantly all kinds of cancer. Since we have not found a guaranteed treatment for him to this day, such studies that contribute to the service of the medical field will have adult approval in society. The particle swarm optimization (PSO) algorithm is used for this purpose by evolutionary algorithms. Obtaining good results by using this algorithm, which is considered one of the important evolutionary algorithms due to the importance of its results through the news of a model using the (PSO).

Keywords: Optimization algorithms, medical data, PSO algorithm

Introduction:

The algorithms are the evolutionary quantum central core in solving improvement problems to reach the optimal solutions or close to that, you may or may not find them at all. The relatively high computational requirements of evolutionary algorithms that we must take into account are largely due to the determination of the fit. This complexity can be reduced by approximating the fitness. [1]

The analysis of patient data or analysis in general for any problem, whether that problem is health, academic or other problems of improvement. We will take a data set of all types of cancer patients and we will try to reach good results. The analysis itself with such data is a step in the right direction because of the direct impact of this study on society. Most of the recent studies that specialize in and analyze high-dimensional medical data aim to reduce those dimensions.[2]

Early diagnosis of all types of cancer aims at a vital and important role in the field of medicine. It will help the doctor and the patient alike as a financial cost second and first to treat the disease before it enters the body of the

person who will be discovered early. There are many classification techniques in which it is not easy to distinguish relevant traits from others, i.e. useless ones, and through the paper, those irrelevant traits must be eliminated. [3]

A sample is taken randomly population (a subset) of those data that reach ten thousand features or more types of cancer. EAs are one of several swarm intelligence-based nature-inspired algorithms, such as the genetic algorithm and particle swarm optimization. It has been applied in a variety of fields, including model identification problems, scheduling problems, parameter optimization problems, the traveling salesman problem, the unit commitment problem, the distribution problem, and the controller problem, due to its characteristics, which include a simple concept, reduced parameters, powerful optimal performance, fast calculation speed, and easy realization. [4]

Problem Statement and contribution :

What is artificial intelligence is the ability of a computer or machine mind to learn from the surrounding environment without being preprogrammed literally, and here it may come to the minds of many, is this possible? and how?

The answer is possible. Yes, it is certainly possible. As how it is through two main components: data and algorithms. [5]

Let's take a simple example, which is the game of chess. When programming the game, every movement made by the game machines must be taken into consideration in each situation. As for artificial intelligence, the solution to the problem is different by using what was mentioned of algorithms and data.

The data is a large group of images of different movements and poses accompanied by the ideal movement, and here we must focus on the word ideal movement.

Then the algorithm comes to this data. The algorithm tries to infer or elicit a general relationship or pattern through continuous training on the data and after obtaining this relationship (similar to a mathematical law) then we say that the robot brain has learned and this is called the training period for the robot brain or its development Then this relationship is used for prediction. [6]

The ability of machine learning from the surrounding environment (the data) and here in the example that was taken into account are pictures of the different chess positions without writing that programmatically in advance literally.

Algorithms came to solve problems that are difficult to solve by traditional methods, such as the traveling salesman problem. If we have such huge, high-dimensional data and have a direct impact on society, we must contribute to solving those problems by using algorithms that have the ability to deal with big problems to get good results using particle swarm optimization (P S O). [7]

In the last decade, genetic algorithms have become more widely used in engineering since they are seen as a tool for optimizing engineering design. The use of evolutionary algorithms to tackle difficult optimization issues and manage uncertainty in common industrial scenarios is discussed in this paper. In this context, the GA is used as an optimization technique for decision making in the Supply Chain planning problem under uncertainty to establish the optimal production, inventory level, and distribution. A MILP stochastic approach was addressed using a basic Matlab genetic algorithm, yielding better results. [8]

Objective and contribution:

For the purpose of early cancer detection, an accurate diagnosis must be made, which is not easy, because genes contain tens of thousands of features for each sample, and these features are not all relevant, and the reason for the problem's complexity is the unimportant features that must be eliminated in order to determine the important features, which will contribute to the disease's early detection (cancer). Data with high-dimensional features has a detrimental impact on the diagnosis mechanism; thus, try to decrease these undesired features by employing metaheuristic algorithms, such as those used in this research (Particle Swarm Optimization) They have certain qualities that set them apart, and they have solved numerous selection difficulties.

Objective:

In this study, a novel features selection model based on evolutionary algorithms PSO has been built, with the goal of reducing unnecessary characteristics in high-dimensional biological data for enhanced diagnosis accuracy.

The proposed model is used in conjunction with the PSO evolution method to optimize feature selection and obtain the best results during the search phase. The proposed strategy draws inspiration from a variety of sources in order to produce innovative solutions that boost research performance.

Goals of Expected:

1- The most important challenge is the high-dimensional medical data, after reducing it, we will reach the correct diagnosis at high rates.

2- Using the proposed model, we get the optimal or close to it solution by choosing a population at random.
3- The PSO is an algorithm inspired by nature or the nature of living organisms. It chooses the best and it is very possible to reach the optimal solution.

Related Work:

To optimize feature selection (FS) techniques, several studies have been proposed. Most of these studies applied standard evolutionary algorithms such as particle swarm optimization (PSO) and bat algorithm (BA). Researchers in the area of optimization FS often applied the original version of evolutionary algorithms to select optimal features without improving the vulnerabilities in those algorithms.[9]

Amrita et al. [10] proposed a hybrid model based on BA to optimize paraments of support vector machine (SVM) and selected optimal features. Each bat vector in the optimization pool consists of two parts: the first two positions allocated to SVM parameters and the rest of the vet represents sent the mask to select features. The proposed system optimizes simultaneously the SVM parameters and features. Applying the evolutionary to optimize more than one part of the system increases the bearable randomization and unstable in final results.[11]

Taha et al. [15] applied naïve Bayes (NB) to guide BA in selecting optimal subset features. They proposed decreeing bat velocity when the different past and current position is negative. NB sets bat velocity to adapt the number of features with the current search process. The feature that corresponds 1 in bat vector is selected and 0 is omitted.[12]

Taha et al. [13] proposed the binary version of particle swarm optimization (BPSO) for optimizing the features selection. The proposed system applies k-nearest neighbors (KNN) as a classifier of the new query. It improves the performance of KNN without developing the PSO.

Dhrif et al. [14] developed PSO by modifying the coefficient parameters (c_1, c_2) and weight inertia simultaneously with search progress. The sigmoid function is applied to enhance these parameters. The authors introduced a new encoding technique that keeps the dynamics of the PSO particles by preserving the continuous behavior of their search process.

Search space in [15] is limited to either 0 or 1 (binary search space) this restricted the search operations to a small range. Generally, the probability stuck at the local optimum is increased when metaheuristic algorithms search in binary search space [16].

Ray et al. [17] proposed multi-operator differential evolution (DE) for feature selection. They extended search space by using continuous search space rather than binary. Authors set the threshold for feature selection to select features that correspond greater than the threshold.

Methodology :

Evolutionary algorithms (using PSO) is one of the most interested and widely used bio-inspired evolutionary in computer engineering fields. PSO is inspired by the social behavior of bird flocking, After the success of using this algorithm to solve single and multi-objective optimization problems, it becomes popular to solve many different issues, and problems. were proposed this Model using (PSO) algorithm to solve medical high dimensional using PSOs parallel.

The proposed model begins by setting PSO starting settings.

PARAMETER SETTING

The parameters of the algorithm used in the experiments, performance, and test problems are explained in this chapter. In this paper, will know how to use it in the experiments The algorithm PSO, as a model.

Data visualizations that are interactive and extremely configurable. First of all, will answer about important ask that, ***Why are we using MATLAB to implement our work and create an environment?***

Engineers and scientists use MATLAB to organize, clean, and analyze complex data sets from diverse fields such as climatology, predictive maintenance, medical research, and finance MATLAB provides:

1- Scientific and engineering data datatypes and preprocessing capabilities.
2- Interactive data visualizations with a wide range of options.
3- Thousands of prebuilt statistical, machine learning, and signal processing routines.
4- Professionally written and extensive documentation.
5- Professionally written and extensive documentation.
6- Data analysis is expanded without requiring significant programming changes.

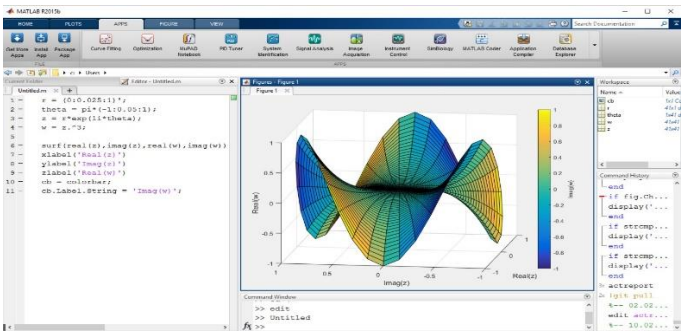


Figure (1): Matlab environment for the proposed model.

MHFS Parameters

To implement a fair experiments framework, the algorithm's parameters were chosen to be the same for every test instance. The proposed MHFS begins by setting both PSO starting settings. It has a random value with zero velocity in its population. In PSO, the population size is the same. The population's initial best local optimum is chosen at random. As a result, in the first stage, the suggested model does not exchange the best solutions.

Table (1) results of the experiment for biomedical data(experiment1)[21]

NO.	Name of dataset	Number of Classes	Number of Patients	Number of Attributes	Accuracy C.F/O
1	**NCI**	7	58	5244	633
2	**Brain_tumor1**	6	87	5920	408
3	**Adenocarcinoma**	4	71	9868	556
4	**Lymph cancer**	5	66	4025	305
5	**Leuk_cancer**	5	41	3050	372

Table 2: results of the experiment for biomedical data(experiment1) [34]

NO.	Name of dataset	Number of Classes	Number of Patients	Number of Attributes	Accuracy CF/O
1	**14_tumors**	9	124	15,010	1089
2	**9_tumors**	8	55	5726	833
3	**Brain_tumor2**	3	42	10,367	741
4	**11_tumors**	7	102	12,533	860
5	**Brain_tumor2**	2	39	5597	287

Obstacles with Implementation

1. ***Local_ Optimum Stagnancy***:

PSOs often identify the initial local_ optimum solution and then lose interest in improving their search position. In an evolutionary algorithm, the only way to produce new candidate solutions is through weak exploration. As a result, tremendous suffering is required to overcome the local optimum and reach the best one.

Inaccurate when dealing with a large number of dimensions:

When dealing with the high dimensional search problem, the PSO suffers greatly. They usually fall into local optimums and have a slow convergence rate as the search progresses.

2. ***PERMANENT OUTCOMES:***

PSO solution engines are unstable, and their output solutions are extremely different, especially when dealing with high-dimensional problems.

The millions of features recorded for each sample present a hurdle for certain gene expression profile categorization approaches.

Mechanism for Searching (Proposed Model)

This phase enables the search of the algorithm (PSO). During the search process, The PSO updates the position of its particle and changes the position of the PSOS= PS ii=1...m, particle swarm populations, respectively, PS I being the solution vectors, and f being the fitness(cost) function. algorithm demonstrates search methodologies that have been developed.

A mixture of solutions

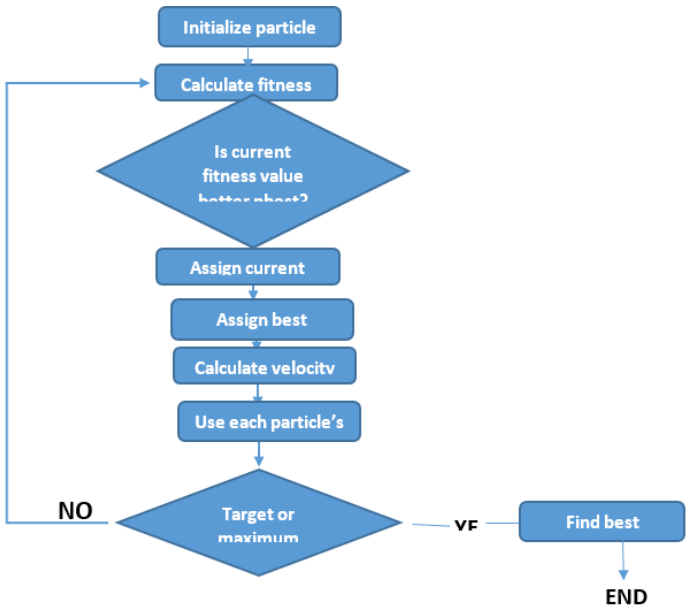
- *exceeding the max iteration,*
- *making no progress search,*
- *satisfying the equbment requirement.*

The proposed model, which is supported by PSO, selects the best option in this part. The optimal solution is chosen using two-equation

$$v_i^k = wv_i^k + c_1 r_1 \left(\text{pbest}_i^k - x_i^k\right) + c_2 r_2 \left(\text{gbest}^k - x_i^k\right)$$
$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

When the maximum iteration is reached, the suggested system comes to a halt

PSO algorithm Implementation



Proposed Model Implementation

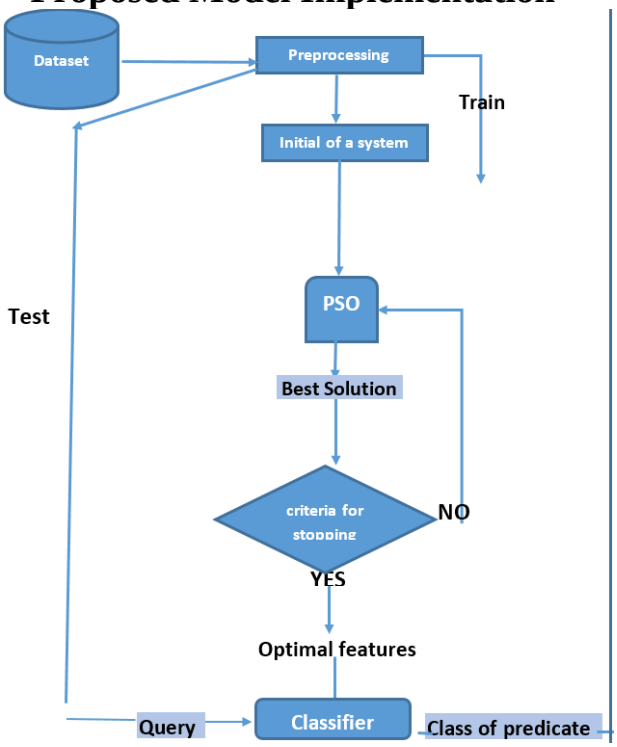


Figure (3.10) Implementation of Proposed PSO Model.

Figure 3.10 shows the proposed model for a paper that contains two parts train then enter in test by using algorithms PSO to obtain the best solution after taking the results of PSO.

3-8 Experiments and discussions

The datasets that were utilized to evaluate the proposed system are given in this section, as well as the findings obtained by the proposed model that usıng PSO Algorıthm.

3-8.1 PARAMETER SETTING **DATASET**

The proposed approach model is to test on ten different medical datasets. They came from

Table (3.3) parameters of the dataset.[21]

Auto NO.	Name of the dataset	No. Item	No. Patients	No. Attributes	F.C/O
1	DLBCL1	2	77	5469	142
2	Breast_3	3	95	4869	154
3	Prostate_cancer	2	38	3051	161
4	Lymph_cancer	3	62	4026	195
5	Prostate tumor1	2	102	10510	206

two different websites. The first is at, which has five medıcal datasets: (, cancer, prostate tumor, 9 tumors, brain tumor1, and DLBCL), and the second is at, which contains the remaining datasets in table 1(Adenocarcinoma1, Lung _cancer, Brain_tumor2,9_tumors), as shown in Table (1,2) respctlly.

The table show complexity of the data measured parameters where C.F/O has a high dimension problem because ***C*** is the number of classes where F is the number of features, and ***O*** is the number of observations. Technical a bigger number of features than the number of observations means the dataset has high complexity.

3.8.2 EXPERIMENTS AND DISCUSSIONS

The experimental evaluation took into account a wide variety of biomedical dataset features numbers, as shown in table1(3.1), in order to demonstrate an accurate assessment of the suggested features selection technique.

The K Nearest Neighbor (KNN) is used as an ***objective function*** to features selected by (PSO), with the nearest neighbor (K) set to 1. The best prediction

accuracy, high persistent outcomes, and better exploration throughout thirty the times run of the algorithm. ***The standard deviation (Std)*** is used to quantify the consistency of the results we can use as a metric for performance. The average prediction accuracy of ***thirty times*** is derived as the accuracy performance of all comparison results algorithms. The exploration rate of the PSO and PSO is computed using the ***average standard deviation(ASD)*** for the number of features exported during the search process.

Table (3.4) the experimental results of MHFS PSO-BA, BA, and PSO accuracy and number of features across thirty runs.

Sq.	Dataset	PSO Proposed System	
		accuracy	Number of feature
1.	DLBCL_cancer	**99.74**	359
2.	Breast3_cancer	78.19	97
3.	Prostate_cancer	99	159
4.	Lymph_vancer	93.88	160
5.	Prostate_tumor1	**94.78**	231

Metrics and Performance:

When we focus on the Tables show the performances of the proposed algorithm PSO Model against the other (PSO) compared algorithms for the ten medical data sets respectively. For each experiment, the best result is represented in bold. The values in the table represent the average values of the independent runs for each instance of the experiments. As we can see from the tables, the proposed algorithm that uses our new model mechanism outperforms the well-known PSO algorithm in most cases.

The values in the table were calculated using the standard deviation coefficient of the given mathematical law described, or using the Excel program by writing one function, we find the (STD).

Through the existing values, we notice a clear superiority of the proposed system over the two algorithms that were used in the research, namely, the particle swarm optimiz ation algorithm and the bat algorithm.

The symbol (Bold) means that the best algorithm significantly outperforms the other algorithm result in the same row and test instance where the significance level as a statistical test to investigate our results more accurately.

work in this project using the two algorithms (the particle swarm optimization algorithm), each of which gives specific results represented in the number of features and the percentage of their work accuracy, experimentally, the number of features is much less than the features of the original dataset, for example using the bat algorithm using the cancer dataset that contains more than 5000 features, only 681 features were selected, with an accuracy rate of 95.1%.

while using the particle swarm optimization algorithm on the same dataset, 595 features were selected with an accuracy rate of 96.1, taking into account that we have not yet used the proposed system So far, using the proposed hybrid system in this project, we have obtained 184 features and an accuracy of 99.5, and we can develop it in the future. To summarize, when compared to previous algorithms, the suggested system (PSO) obtained superior accuracy with a lesser number of selected features and high accuracy in 80% of datasets. It also met a measurable preperformance requirement for the most difficult datasets (high F.C/O).

102x6033 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.8630	-0.8630	-0.4713	-0.8630	-0.8630	-0.8630	-0.8630	4.2954	-0.8630	-0.8630	0.8518	-0.8630
2	-0.8321	-0.8321	-0.8321	-0.8321	-0.8321	-0.8321	-0.8321	4.0169	-0.8321	-0.8321	0.5715	-0.3294
3	-0.7830	-0.7830	-0.5341	-0.7830	-0.7830	-0.7830	-0.7830	5.0668	-0.7830	-0.7830	0.4053	-0.7830
4	-0.7988	-0.5514	-0.4165	-0.7988	-0.7988	-0.7988	-0.5514	4.6856	-0.7988	-0.7988	0.3553	-0.7988
5	-0.8124	-0.8124	-0.8124	-0.8124	-0.8124	-0.8124	-0.8124	3.8432	-0.8124	-0.8124	0.6219	-0.8124
6	-0.7966	-0.7966	-0.6205	-0.7966	-0.7966	-0.7966	-0.7966	4.4808	-0.7966	-0.7966	0.5666	-0.7966
7	-0.8155	-0.8155	-0.5637	-0.8155	-0.8155	-0.8155	-0.8155	4.6442	-0.8155	-0.8155	0.5621	-0.8155
8	-0.8625	-0.8625	-0.2861	-0.8625	-0.6052	-0.8625	-0.8625	4.8692	-0.8625	-0.8625	1.1264	-0.8625
9	-0.6900	-0.6900	-0.6900	-0.6900	-0.5999	-0.6900	-0.6900	4.4042	-0.6900	-0.6900	0.6888	-0.6900
10	-0.7361	-0.7361	-0.7361	-0.7361	-0.4849	-0.7361	-0.7361	4.1721	-0.7361	-0.7361	0.8785	-0.7361
11	-0.7143	-1.2069	-1.2069	-1.2069	0.3746	1.1062	0.3356	3.8374	0.6620	-1.2069	-0.2321	-0.8671
12	-0.7177	-0.7177	-0.7177	-0.7177	-0.6270	-0.7177	-0.7177	4.3528	-0.7177	-0.7177	0.2618	-0.7177
13	-0.8878	-0.8878	-0.2516	-0.8878	-0.8878	-0.8878	-0.7934	5.4943	-0.8878	-0.8878	0.7074	0.4354
14	-0.8089	-0.8089	-0.3490	-0.8089	-0.8089	-0.8089	-0.8089	5.0626	-0.8089	-0.8089	0.4446	-0.8089
15	-0.8068	-0.8068	-0.4150	-0.8068	-0.8068	-0.8068	-0.8068	4.6373	-0.8068	-0.8068	0.5328	-0.8068
16	-0.8826	-0.8826	-0.0630	-0.8826	-0.8826	-0.0211	-0.7032	4.7416	-0.7888	-0.8826	1.0037	-0.4201
17	-0.8975	-0.8975	-0.7173	-0.8975	-0.8975	-0.8975	-0.8975	4.9286	-0.8975	-0.8975	0.3405	-0.8975
18	-0.7981	-0.7981	-0.5436	-0.7981	-0.7981	-0.7981	-0.7981	4.7735	-0.7981	-0.7981	0.6609	-0.7981
19	-0.9526	-0.3284	0.0828	-0.9526	-0.9526	0.4429	-0.9526	4.8144	-0.6974	-0.9526	0.0828	0.5314
20	-0.9432	-0.0808	-0.6847	-0.9432	0.1056	-0.7636	-0.9432	5.0873	-0.9432	-0.9432	1.4540	-0.2123
21	-0.8303	-0.8303	-0.8303	-0.8303	-0.8303	-0.8303	-0.8303	3.9895	-0.8303	-0.8303	0.9706	-0.8303
22	-0.3063	-1.1776	-1.1776	-1.0907	0.1310	-0.0617	-0.5923	4.4823	-0.3421	-1.1776	-0.1758	-0.2719
23	-1.0386	-0.7074	0.0749	-1.0386	-1.0386	0.2491	-0.3084	4.7468	-0.5164	-1.0386	-0.1368	-1.0386
24	-0.8523	-0.8523	-0.8523	-0.8523	-0.8523	-0.5241	-0.3939	4.7822	-0.8523	-0.8523	0.5475	-0.8523

Figure(3.11) Sample of Dataset of prostate

This dataset(prostate) that consist of high dimensional biomedical data must be reduced for this hug data using algorithms PSO, In addition to the proposed system, it consists of 3051 features, 38 samples, and 2 to reduce data to easy operation complexity and diagnosis patient due to the high cost and long time obtain on to 30 features from 9869 features. the degree of complexity of high dimensional biomedical data results from a number of features multiplicative with the number of classes all these operations divided by the number of samples (how many patient assessment diagnoses) and classes are personal that is done testing by using two algorithms particle swarm optimization and Bat algorithm progress operation in parallel to find the best solution after entering dataset for the proposed system, to train any

machine learning divided data into two parts train in rate 80% approximately and 20% tested.

system old with a proposed system that consists of algorithm, stop criteria either yes go to the final result or no return to beginner system (system initialization), after obtain on the optimal features to be accessible to the end system model, this model depends on an accuracy number of features of (PSO), PSO works PSO algorithm, it satisfied a markable performance over highest datasets (high C.F/O) needs to experiment more number of features to enhances the proposed system to provide an efficient exploration compared with other algorithms.

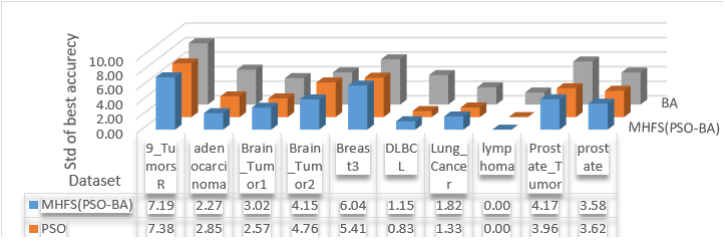


Figure 3.12 performance for 30 times using standard deviation

The values in the figure (3.12) were calculated using the standard deviation coefficient of the given mathematical law described below or using the Excel program by writing one function, we find the (STD).

Through the existing values, we notice a clear superiority of the proposed system over the algorithm that was used in the research, namely, the particle swarm optimization algorithm and the bat algorithm.

The best in the proposed system, the less, the better thirty times the implementation, and this was done 30 times because work is evolutionary and means randomness, and maybe someone asks what is the evidence for the validity of the work as a whole because randomness depends on the principle of luck, so the answer is the number of times of implementation by 30 Once and each of them is repeated twenty times, so 20 in 30 is equal to 600 times. This process is repeated on the dataset that was used during the research.

It is not said that all these 600 times came by chance, if the implementation was once or twice, the data could have come by chance.

Finally, the higher the standard deviation, the results will be farther apart, and the less it will be, the results will be close. For example, if the results are as follows (90, 60, 90, 30, 40), then it is normal that the standard deviation will be higher. But if the results are close (90, 91, 92, 90, 91), the standard deviation will be much less, which is the best, and the proposed hybrid system will be better than the two previous traditional algorithms.

We can compute(S.T.D) from excel function (=stdv.s) and then the exchanging best solution enhances the proposed system to provide an efficient exploration with of algorithm (PSO).

Figure 5 shows the exploration rate of the proposed system, PSO.

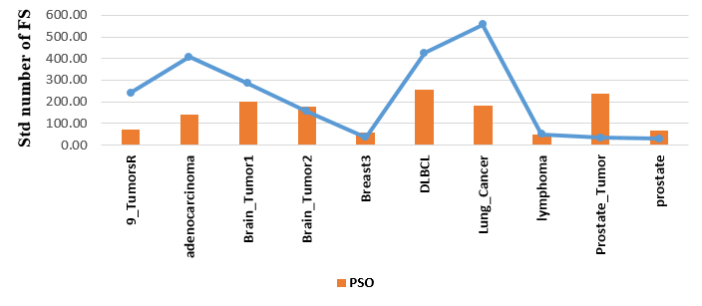


Figure 3.13 average of search model, rate PSO and BA

This process is repeated on the dataset that was used during the research.

It is not said that all these 600 times came by chance, if the implementation was once or twice, the data could have come by chance.

Conclusion and future work:

Finally, due to weaknesses in their exploration, EAs algorithms frequently experience stagnation during the search process.

1- Result, improving the exploration of these algorithms requires raising the diversity solution in search progress.

2- Suggested a new search methodology based on multiple EAs (PSO) in this study for project,

3- Greatly increased the exploration rate. Our suggested improved results by using the PSO to search for optimal solutions the best solutions algorithms as the search progressed.

4- The proposed model is compared with PSO to select the best group of features that enhance the classification accuracy of the biomedical dataset.

5- The suggested has outperformed other models in experiments and has been shown to be a viable optimization alternative model for feature selection. For future work.

6- A proactive random model would be applied to adaptive the variable search

Will have achieved the goal of the research, which will help in diagnosing cancer, which is one of the diseases that contains high-dimensional data and is difficult to diagnose by ordinary traditional methods, so we used metaphysical methods to solve such complex problems. A set of experiments

were performed to investigate the performance of the model proposed. In these experiments, the performance of our model was compared to the PSO,

References

1- Mitchell, M. (2019). Artificial intelligence: A guide for thinking humans. Penguin UK.
2- Thiagarajan, J. J., Thopalli, K., Rajan, D., & Turaga, P. (2022). Training calibration-based counterfactual explainers for deep learning models in medical image analysis. Scientific Reports, 12(1), 1-15.
3- Ebadinezhad, S. (2020). DEACO: Adopting dynamic evaporation strategy to enhance ACO algorithm for the traveling salesman problem. Engineering Applications of Artificial Intelligence, 92, 103649.
4- Saghaei, M., Ghaderi, H., & Soleimani, H. (2020). Design and optimization of biomass electricity supply chain with uncertainty in material quality, availability and market demand. Energy, 197, 117165.
5- A. M. Taha, A. Mustapha, and S. Der Chen, "Naive Bayes-guided bat algorithm for feature selection," Sci. World J., vol. 2013, 2013.
6- Amrita and P. Ahmed, "A hybrid-based feature selection approach for IDS," Lect. Notes Electr. Eng., vol. 284 LNEE, pp. 195–211, 2014.
7- T. Ray, R. Sarker, and X. Li, "Preface," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9592, p. v, 2016.
8- G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Comput. Electr. Eng., vol. 40, no. 1, pp. 16–28, 2014.
9- H. Zeng and H. J. Trussell, "Feature selection using," IAENG Int. J. Comput. Sci., vol. 33, no. February, pp. 997–1000, 2006.
10- C. H. Park and S. B. Kim, "Sequential random k-nearest neighbor feature selection for high-dimensional data," Expert Syst. Appl., vol. 42, no. 5, pp. 2336–2342, 2015.
11- Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," Knowledge-Based Syst., vol. 64, pp. 22–31, 2014.
12. S. Member and S. Member, "A Mean-Variance Optimization Algorithm," pp. 1–6, 2010.
13. A. Kaveh and N. Farhoudi, "A new optimization method: Dolphin echolocation," Adv. Eng. Softw., vol. 59, pp. 53–70, 2013.
12- P. Hansen and N. Mladenović, "Variable neighborhood search," Handb. Heuristics, vol. 1–2, no. 1, pp. 759–787, 2018.

13.R. Yu, X. An, B. Jin, J. Shi, O. A. Move, and Y. Liu, "Particle classification optimization-based BP network for telecommunication customer churn prediction," Neural Comput. Appl., vol. 29, no. 3, pp. 707–720, 2018.
14. M. Z. Zakaria, S. Mutalib, S. A. Rahman, S. J. Elias, and A. Z. Shahuddin, "Solving RFID mobile reader path problem with optimization algorithms," Indones. J. Electr. Eng. Comput. Sci., vol. 13, no. 3, pp. 1110–1116, 2019.
15. [7] N. Z. Mohd Ali, I. Musirin, and H. Mohamad, "Clonal evolutionary particle swarm optimization for congestion management and compensation scheme in power system," Indones. J. Electr. Eng. Comput. Sci., vol. 16, no. 2, pp. 591–598, 2019.
16. A. H. Jabor and A. H. Ali, "Dual Heuristic Feature Selection Based on Genetic Algorithm and Binary Particle Swarm Optimization," J. Univ. BABYLON pure Appl. Sci., vol. 27, no. 1, pp. 171–183, 2019.
17. A. O. Topal and O. Altun, "A novel meta-heuristic algorithm: Dynamic Virtual Bats Algorithm," Inf. Sci. (Ny)., vol. 354, pp. 222–235,2016

الباحث صفاء جبر محمد
المديرية العامة لتربية الديوانية
Safaa.Fuzzy@gmail.com
+9647718884035

مستخلص البحث:
تعتبر مشاكل الأبعاد العالية للبيانات الطبية أهم شيء في أي تحليل لمجموعة البيانات ، خاصة تلك التي تختص بالبيانات الطبية عالية الأبعاد ، حيث لا يمكنك تحديد الميزة التي تقع ضمن آلاف الميزات الموجودة في كل كروموسوم و يوجد في كل منها آلاف الجينات.
الغرض الرئيسي من البحث يمكن تلخصيه بالنقاط الاتية :-
1- التشخيص الصحيح المبكر لمرض الكانسر بنسبة عالية جدا تصل الى 99% .
2- التقليل من التكاليف الباهضة التي يتحملها المريض اثناء الفحوصات العادية .
3- مساعدة الطبيب باستخدام تكنولوجيا متقدمة جدا لا تقبل الملل او التعب وبالتالي يمكن اعتماد النتائج .
4- السبب الرئيسي لاختيار مرض الكانسر كونه من الامراض التي لم يتم التوصل لعلاجه بشكل نهائي حتى اليوم .
5- من جهة اخرى فأن الدراسة التي لم تتأثر بالمجتمع تعتبر ركيكة نوعا ما وان مثل هذه الدراسات تعتبر تقدم كبير في مجال الطب .
بدايةً تم التعامل مع مجموعة البيانات التي تخص الكانسر المرفوعة من قبل باحثين سابقين من خلال اللغة البرمجية الشهيرة الماتلاب 2021 بخطوات متسلسلة كالاتي
Preprocessing ….. initialization ….. setting parameter of PSO algorithm ….
Population size … make the population .