Journal of Computing and Applied Informatics (JoCAI) Vol. 04, No. 1, 2020 | 1 – 15

talenta PUBLISHER

DATA SCIENCE

Journal of Computing and Applied Informatics

Improving KNN by Gases Brownian Motion Optimization Algorithm to Breast Cancer Detection

Majid Abdolrazzagh-Nezhad[1*], Shokooh Pour Mahyabadi[2], and Ali Ebrahimpoor[3]

[1,3]*Department of Computer Engineering, Faculty of Engineering, Bozorgmehr University of Qaenat, Qaen, Iran*

[2]*Department of Computer Engineering, Birjand Branch, Islamic Azad University, Birjand, Iran*

Abstract. In the last decade, the application of information technology and artificial intelligence algorithms are widely developed in collecting information of cancer patients and detecting them based on proposing various detection algorithms. The K-Nearest-Neighbor classification algorithm (KNN) is one of the most popular of detection algorithms, which has two challenges in determining the value of k and the volume of computations proportional to the size of the data and sample selected for training. In this paper, the Gaussian Brownian Motion Optimization (GBMO) algorithm is utilized for improving the KNN performance to breast cancer detection. To achieve to this aim, each gas molecule contains the information such as a selected subset of features to apply the KNN and k value. The GBMO has lower time-complexity order than other algorithms and has also been observed to perform better than other optimization algorithms in other applications. The algorithm and three well-known meta-heuristic algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Imperialist Competitive Algorithm (ICA) have been implemented on five benchmark functions and compared the obtained results. The GBMO+KNN performed on three benchmark datasets of breast cancer from UCI and the obtained results are compared with other existing cancer detection algorithms. These comparisons show significantly improves this classification accuracy with the proposed detection algorithm.

Keyword: Breast Cancer Detection, Classification, K-Nearest-Neighbor Algorithm, Feature Selection, Gases Brownian Motion Optimization.

Received 20 September 2019 | Revised 25 December 2019 | Accepted 29 January 2020

1. Introduction

Cancer is one of the most common causes of death today. According to the National Cancer Institute, prostate, breast, lung and intestine are the most common sites of cancer, respectively. The traditional treatments of cancers such as chemotherapy and radiation are problematic. Application of information technology in health issue is one of the important effects of technology or perhaps life-changing by detection will get faster and better, and medicine becomes

*Corresponding author at: Department of Computer Engineering, Faculty of Engineering, Bozorgmehr University of Qaenat, Qaen, Iran

E-mail address: abdolrazzagh@buqaen.ac.ir

ISSN: 2580-6769 | e-ISSN: 2580-829X | DOI : https://doi.org/10.32734/jocai.v4.i1-3619
Journal Homepage: https://jocai.usu.ac.id

personalized. So, the technology has become more justifiable and the disease could be detected before its serious signs are seen. Actually every one could be become his/her own doctor. Various classification algorithms have been investigated and developed to cancer detection such as Artificial Neural Networks (ANN) [1], Decision Trees (DT) [2], Bayesian Networks (BN) [3], Naïve Bayesian (NB) [4], Logistic Regression (LR) [5], Support Vector Machine (SVM) [6], Hyper-Plane Classifiers (HPC) [7] and K-Nearest-Neighbor (KNN) [8].

The mentioned algorithms need to be improved by meta-heuristic algorithms, because they have the challenges which involved in determining the optimal value of their parameters. Wang et al. combined GA and SVM to feature selection and improved parameters for Brain-Computer Interface Patients (BCI) application [9]. The BCI with the syndrome can communicate with the world by using BCI application. The main component of this system is the extraction feature and pattern recognition. Abdolrazzagh-Nezhad and Radgohar designed a hybridization of SVM and Cultural Algorithm (CA) to cancer detection [10]. Jalayeri and Abdolrazzagh-Nezhad proposed HPC to fill up the challenges of SVM and optimized the HPC by utilized Chemical Reaction Optimization (CRO) [7]. Wu et al. designed a hybridization of SVM and Memetic Algorithm (MA), which called MSVM, to classification large imbalanced data [11]. In the hybridization, the MA is utilized as a heuristic framework to classify large imbalanced data. Due to the high performance of SVM in balanced binary classification, SVM is combined with MA to improve SVM classification accuracy.

Kun Jeng et al. [12] developed a hybrid method. They combined the Borderline Synthetic Minority Oversampling (BSM) and Artificial Immune Detection (AIRS) method as a global optimization explorer(seeker), along with the nearest neighbor algorithm, which is used as a classifier. This method was called BSMAIRS. Among all meta-heuristic algorithms, the AIRS Artificial immune detection system is a popular algorithm which is widely used in medical classification problem inspired by the immune algorithm [14 ,13]. Saeedi et al. [15] developed AIRS to identify diabetes. Poulat et al. [13] utilize AIRS to breast cancer detection and liver disease.

In 2014, Hui Lin Cheng et al. [14] proposed the optimization algorithm of ad hoc parameters to optimize parameters and SVM feature selection simultaneously, named their method PTVPSO - SVM. This method was implemented in a parallel environment by utilizing a parallel virtual machine. In the proposed method, a weighted function is adopted to design the objective function of the PSO, which respects the accuracy average of the SVM classifier, the number of machine vectors and selected features. Moreover, the mutation operator is introduced to resolve the early convergence problem of the PSO algorithm to increase the PSO algorithm performance. In sum,

the improved PSO binary algorithm is very efficient in enhancing the PSO algorithm performance in feature selection.

Huang et al. [16] also proposed a distributed PSO - SVM hybrid system with feature selection and parameter optimization. The Support Vector Machine is a popular classification method with many different applications. The kernel parameter tuning in the SVM training procedure, along with feature selection, significantly encompasses the precision of classification. This study determines the parameter values while determining a subset of features without any reduction of classification accuracy simultaneously. The particle swarm optimization-based method to determine the parameters and selection of the SVM feature was called the PSO – SVM.

Soudhid et al. [17] proposed a novel method based on the combination of the firefly algorithm and SVM to predict malaria. Axia Yong et al. [18] proposed a combination of SVM methods and particle swarm optimization and cuckoo search algorithm (CS) for illnesses detection. It consists of two phases: the first phase, the CS-based approach to optimize the SVM parameters to find initial appreciate parameters for the kernel function; the second phase is the PSO to remain training the SVM and finding the SVM parameters. A hybrid meta-heuristic method for medical data classification (H colonies) was presented in 2014 by Saber al-Mahideb and et al. [19]. The hybrid ant - bee colony consists of two phases: an optimization phase of the ant colony (ACO) and an optimization phase of the bee colony (ABC). The ABC supply source becomes a decision - making a list that is constructed during the ACO stage by utilizing different subsets of training data. The ABC task is to optimize the obtained lists.

According to the literature reviewed and the challenges of the KNN classification algorithm in this paper, for the first time, it is attempted to utilize the Gases Brownian Motion algorithm to solve the KNN problem. For this purpose, we attempted to choose the most effective features among the set of features of the cancerous data, consequently reducing the volume of computation of KNN by decreasing the data size. Subsequently, by determining the optimal k-value for a subset of the selected features, the accuracy of the cancer detection could be improved cancer detection accuracy. These objectives are composed by designing the structure of a molecule in terms of one bit for a value of K and assuming n features, the number of n binary bits representing the subset of the selected attributes of n main features. The results of this algorithm are compared with genetic algorithms, particle swarm, and Imperialist Competitive Algorithms.

The organization of the paper is as follows. In Section 2, the detection phases of cancer are described as the research issue. Subsequently, in Section 3, the process of the Gases Brownian Motion optimization algorithm is stated, and in Section 4, the results of the implementation are compared with the well-known optimization algorithms. Finally, in Section 5, the conclusion and summarization of the paper are presented.

2. Cancer Detection Problem Description

The detection problem of cancer diseases is the classification problem that is related to the supervised learning machine problems [20]. Medical databases where each line (entity) defines a patient and each column (attribute) characterize patient descriptors and clinical and physical test results, and each entity is identified with a class labeled infected or not infected with specific cancer/disease. This problem attempts to identify and construct models by utilizing classification techniques to identify entities with no class label (infected or lack of a specific disease) [21]. In this section, we describe four main steps in the disease detection problem such as data preprocessing, data segmentation, classification and model construction, and finally will compute the accuracy of classification and model evaluation.

2.1 Preprocessing

Data preprocessing and data preparation is the most important step in data mining projects. In the preparation or preprocessing data, first, it should be comprehensive the data purpose and application of data to increase reliability in data mining and consequently the speed of the work is increased. The preprocessing operation, which is necessarily required in this research, is to fill the missing data. One of the missing values management techniques is the use of Mean and Median of features for all samples belonging to the similar class [22]. In some datasets of this study, because of missing values, it is necessary to use a criterion to fill the missing data values. The selection of criteria depends on the asymmetry degree of the probability distribution. When the data distribution is completely symmetric, the mean criterion is appropriate; but when the distribution has skewness, the median is a good criterion. In this study, we selected a suitable central criterion for filling the missing data values depending on the asymmetry of the probability distribution of data.

Another preprocessing requirement in this study is data transformation. At this point, data are transformed into a proper arrangement for data mining. One of the important parts of this stage is data normalization. Normalization involves changing the scale of the data to map data to a small domain such as [1, -1] or [1, 0]. One of the well-known methods for normalization is the Min-Max method. By utilizing the Min-Max normalization, a linear transformation is performed on the original data. Suppose that min_A and max_A are minimum and maximum value of the feature A respectively; by calculating V'_A, the previous values of feature A are mapped to new values in the range [new_max_A ،new_max_A] [22]. V'_A is calculating as follows:

$$V'_A = \frac{v_{i-min_A}}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

At this point, we will obtain the normalized data set by utilizing this equation to our dataset. After applying this method, the existing values of our dataset will be in the range [0, 1].

2.2 K-Nearest Neighbor (KNN) Algorithm

The KNN classification algorithm is a method to classify entities (patients) based on the distance of the training samples from each other and is usually a suitable choice when there is insufficient information about how the data is distributed [23]. The process of the algorithm consists of two main steps. The first step is determining the k nearest neighbor and the second stage is to define the class of the k nearest neighbors. Assume that the training data is defined as follows:

$$D = \{X_1, X_2, \dots, X_n\} \quad (2)$$

Where n is the representation of the number of entities and each entity X_i contains f feature as defined below:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{if}\} \quad (3)$$

The assumed data belonging to Class C are supposed to be different. To determine the X′ data class, firstly the distance of this data to all data available in the D area must be calculated. Then k is determined in this area, which is the smallest distance from X′, and the classes of these data that are in k neighborhood of X′ are specified. Data class of X′ is the similar class that has the greatest frequency among "k - neighborhood" data classes of X'. There are several metrics to calculate the distance between data, but concerning the numerical value of cancer data in this article, the Euclidean metric is utilized in order to compute the distance between the two data X′ and X_i as follows.

$$d(X_i, X_i) = \sqrt{(x_{i1} - x'_1)^2 + \cdots + \left(x_{if} - x'_f\right)^2} \quad (4)$$

Furthermore, in order to implement KNN and data segmentation into two classes of training and testing, the 10-fold method is employed.

3. Gases Brownian Motion Optimization Algorithm

The molecular motion in the gas mode is Brownian motion. For this reason, there is no superiority for molecular speed but it is distributed irregularly in all directions. As a result of the interaction of the molecules, both the direction and velocity of the molecules modify continuously because the gas molecule's velocity are very different. The Brownian nature of the gas molecules motion enables them to quickly travel through space where they are located and occupy the entire volume of space where they are propagated. Based on this idea, the Gases Brownian Motion Optimization (GBMO) Algorithm that is inspired by the nature of the Brownian motion of the molecules in the gas state is proposed for search space of the optimization problems [24]. Based on the available

resources, the performance of this optimization algorithm is significant compared to the previous optimization methods. The steps of this algorithm can be stated as follows.

Step 1 : In the search space, a series of gas molecules are randomly generated. Assuming that there are f attributes for the training data, the structure of a molecule contains f + 1 bits whose first bit corresponds to the k value and the number of the next f bits is related to the selected subset of the training features.

Step 2 : For each molecule, a radial randomly is considered in the interval [0, 1].

Step 3 : Next, the system temperature is initialized to guarantee the convergence of the algorithm. The temperature at the beginning of the search is high and will decrease with time passing. Initially, the search process is searching for a wider range due to the high level of kinetic energy and the speed of the molecules (global search). As time passes, due to the decrease in kinetic energy and the speed of motion of molecules, the search process will have further local search.

Step 4 : molecules Speed and position are calculated and updated according to the following equations.

$$v_i(t+1) = v_i(t) + \sqrt{\frac{3KT}{m}} \quad (5)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (6)$$

Where $v_i(t)$ and $v_i(t+1)$ are the velocity of i molecule at time t and t + 1, and also $x_i(t)$ and $x_i(t+1)$ are the location of i molecule at times t and t + 1.

Step 5 : The Fitness function is called to evaluate the obtained responses. The best responses are kept for comparison with the obtained results. Each time the fitness function is executed, the KNN algorithm is calculated based on the proper value of k and subset of the selected features in the molecule and class accuracy is considered as the amount of fitting value of the molecule. It is worth mentioning that according to the nature of continuous behavior of molecules motion, the values of f + 1 bits of each molecule are discretized before the KNN algorithm recall.

Step 6 : Each molecule tends to vibrate in a certain radius except to moving in different directions. This oscillation creates at the beginning of the search process, generates more local search process and in the last steps it generates more global search. The oscillation of the molecules is modeled by utilizing a chaotic sequence generator.

$$x_i(t+1) = x_i(t) + b - (\frac{a}{2\pi})sin(2\pi x_i(t))mod(1) \quad (7)$$

In the above equation a = 0.5 and b = 0.2 is considered and indicates the location of the molecule's motion. The diagram of the oscillatory motion of the molecules in the gases is shown in Figure 1. The vibration varies for different molecules and in the defined radius.

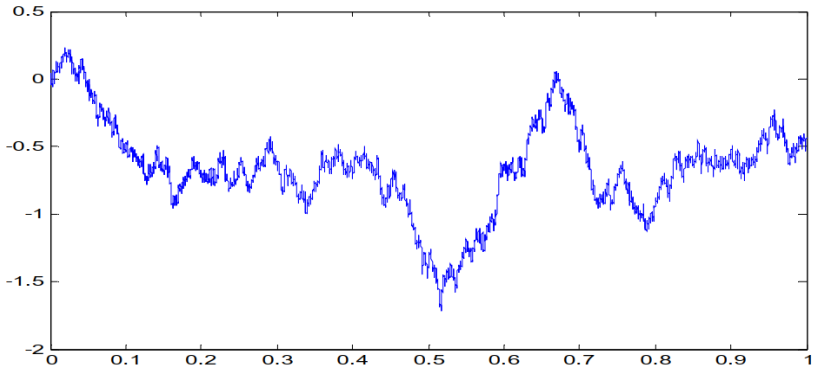

Figure 1. Vibrational motion of molecules with a radius of one [24]

Step 7 : The fitness function is called to evaluate the obtained responses. The best responses are kept for comparison with the obtained results.

Step 8 : Terminate condition of the temperature algorithm is investigated if it reached zero, the algorithm is stopped, and otherwise the search process continues. The global and local search functionality has been implemented in this algorithm, by Brownian motion and molecules oscillation in space, which this motion is shown in Figure 2.

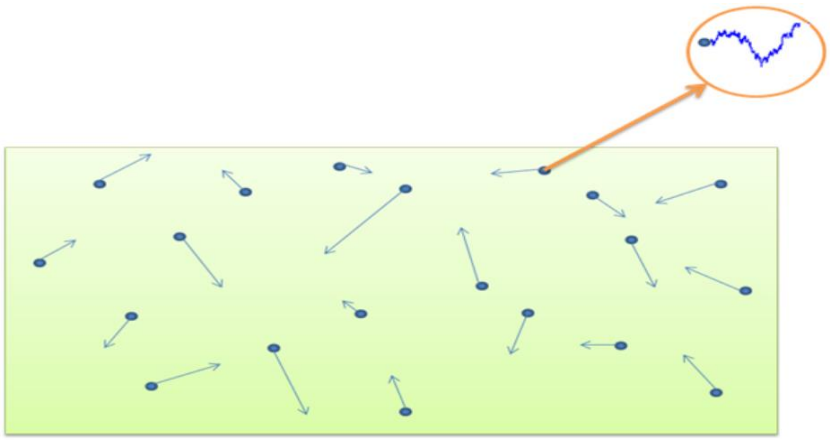

Figure 2. Brownian and vibrational motion of molecules in the problem search space [24]

It is observed by considering Figure 2 that the gas molecules move randomly and irregularly and also vibrate in their place. In Figure 3, the flowchart of the GBMO algorithm is shown.

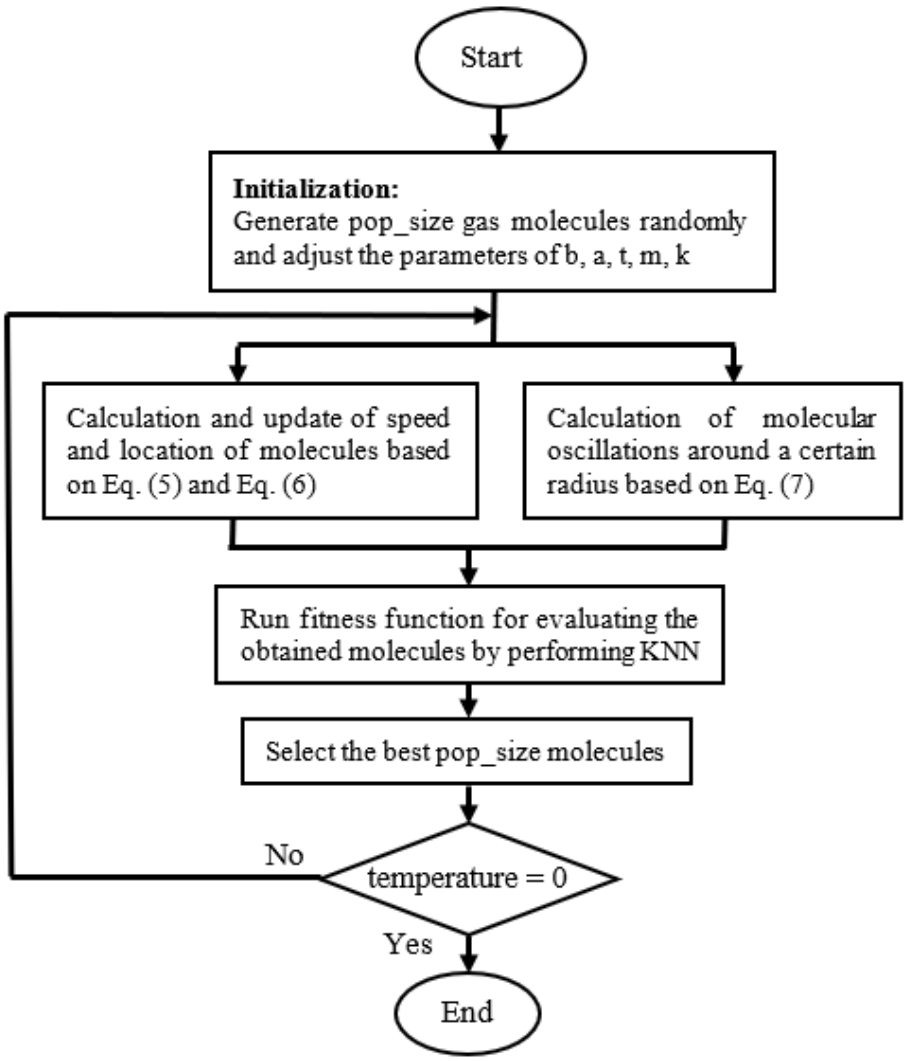

Figure 3. The flowchart of GBMO to improve KNN

4. Experimental Results

For implementation, we used Matlab 2016 software and a system with features of operating system 10 and Intel Core i5 and 6 GB RAM. In order to compare the performance of the GBMO algorithm with the most well-known algorithms such as Genetic Algorithms (GA), Particle swarm (PSO) and Imperialist Competitive Algorithm (ICA), five test functions are considered to be Griewank, Ackley, Sphere, Rosenbrock and Rastrigin, whose details are listed in Table 1. These minimization functions and best of their global response are zero when their dimension is 2 (D = 2).

Table 2 demonstrates the adjusted values for these parameters of algorithms. The parameters adjustments of the algorithms considered in Table 2 are based on values obtained in the previous articles of these algorithms. These algorithms are performed 10 times on the functions of Table 1

and the best detection accuracy, as well as detection accuracy average of cancer obtained, is demonstrated in Tables 3, 4, and 5, respectively.

Table 1. Considered Benchmark functions

Function name	**Equation**
Ackley	$f_1(x) = -20exp\left(-0.2\sqrt{\frac{1}{D}\sum_{i=1}^{D} x_i^2}\right) - exp\left(\frac{1}{D}\sum_{i=1}^{D} cos(2\pi x_i)\right) + 20 + e,$ $-32 \le x_i \le 32$
Rastrigin	$f_2(x) = 10D + \sum_{i=1}^{D}\left(x_i^2 - 10cos(2\pi x_i)\right),\ -5.12 \le x_i \le 5.12$
RosenBrock	$f_3(x) = \sum_{i=1}^{D}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2],\ -2.048 \le x_i \le 2.048$
Sphere	$f_4(x) = \sum_{i=1}^{D} x_i^2,\ -5.12 \le x_i \le 5.12$
Griewank	$f_5(x) = \frac{1}{4000}\sum_{i=1}^{D} x_i^2 - \prod_{i=1}^{D} cos\left(\frac{x_i}{\sqrt{i}}\right) + 1,\ -600 \le x_i \le 600$

Table 2. Parameter adjustments of ICA, GA, PSO and GMBO algorithms

Algorithm	**Parameters**
ICA	NumOfCountries -80, NumOfInitialImperialists - 8, NumOfDecades -1000, Revolution Rate=0.3, AssimilationCoefficient=2, Zeta -0.02
GA	PopSize = 80, MaxGenerations = 1000, CrossPercent=0.5, Mutation Rate 0.2; SelectionMode: Tournment
PSO	C1 = 1.5, C2 = 1.5, B=iter/iter,S=-0.5, ParticleSize = 80, MaxIter- 1000
GBMO	MoleculeSize=80, Temprature - 100, a=0.5, b=0.2

Table 3. Optimal values Mean for functions with D = 10

	GA	**PSO**	**ICA**	**GBMO**
f_1	0.2499	1.14e-17	6.27e-11	4.0e-9
f_2	1.099e-11	0	0	7.98e-9
f_3	48.9673	0.0020	0.4466	8.0096
f_4	0.1016	0.0298	1.93e-7	1.7e-11
f_5	-0.0972	-0.0987	-0.0987	-0.0987

Table 4. Optimal values Mean for functions with D = 20

	GA	**PSO**	**ICA**	**GBMO**
f_1	3.2958	0.0143	0.0732	8.0e-9
f_2	0.6609	5.402e-4	8.403e-4	1.58e-8
f_3	1.32e3	19.517	1.92e3	18.3735
f_4	2.1664	0.2105	0.1205	2.56e-5
f_5	-0.3373	-0.7114	-0.7206	-0.7225

Table 5. Optimal values Mean for functions with D = 30

	GA	**PSO**	**ICA**	**GBMO**
f_1	18.8398	9.0371	28.7678	1.2e-8
f_2	6.3023	0.0015	0.4128	2.38e-8
f_3	9.85e3	28.7533	1.017e6	28.0826
f_4	3.0695	1.0289	19.958	2.56e-5
f_5	-0.3702	-0.3687	-0.8319	-2.3712

By considering the results of Tables 3 and 5, it is observed that by increasing the dimensions of the Benchmark functions, the performance of the GBMO algorithm improves significantly compared to other algorithms. However, in smaller dimensions, the PSO algorithm performs better than the other. In Figs. 4 and 8, the convergence graphs of the best results implemented on the benchmark functions in dimension 30 are illustrated.

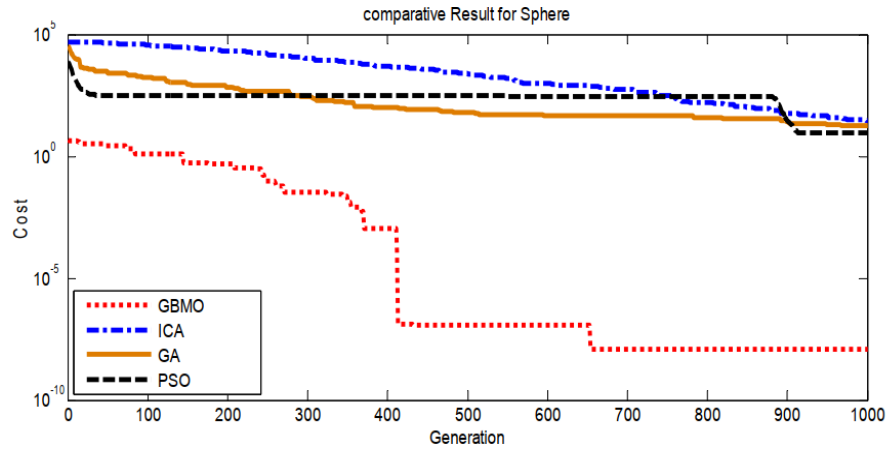

Figure 4. Results Comparison in the Sphere function

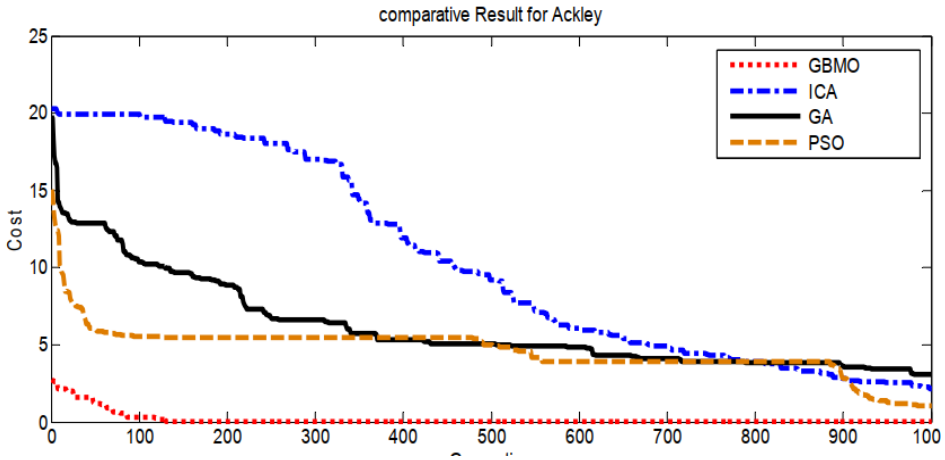

Figure 5. Results Comparison in the Ackley function

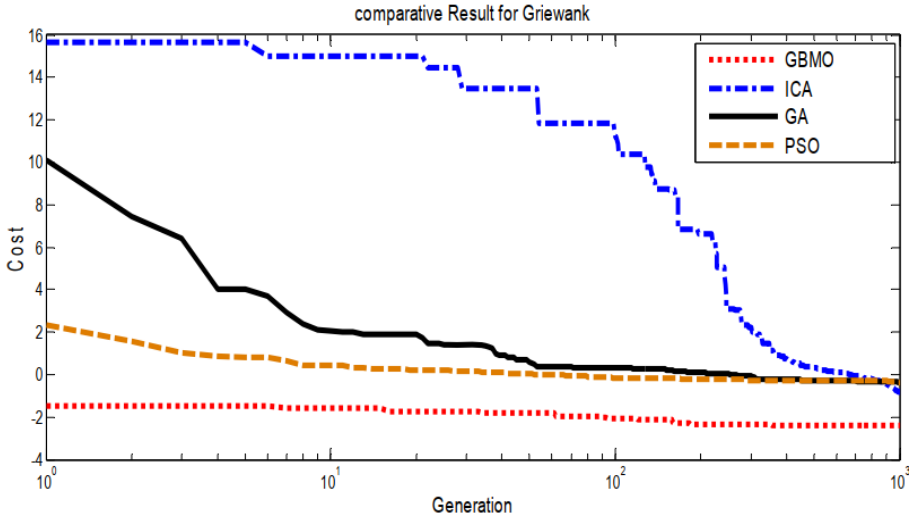

Figure 6. Results Comparison in the Griewank function

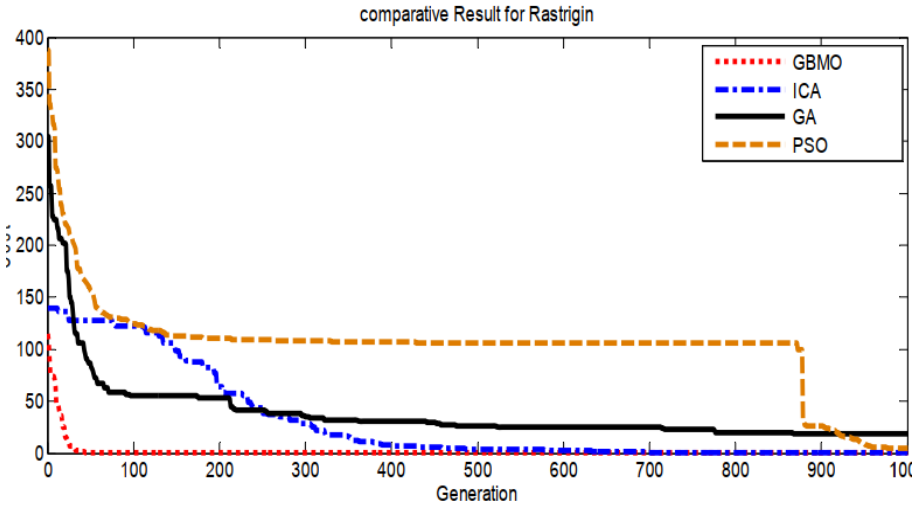

Figure 7. Results Comparison in the Rastrigin function

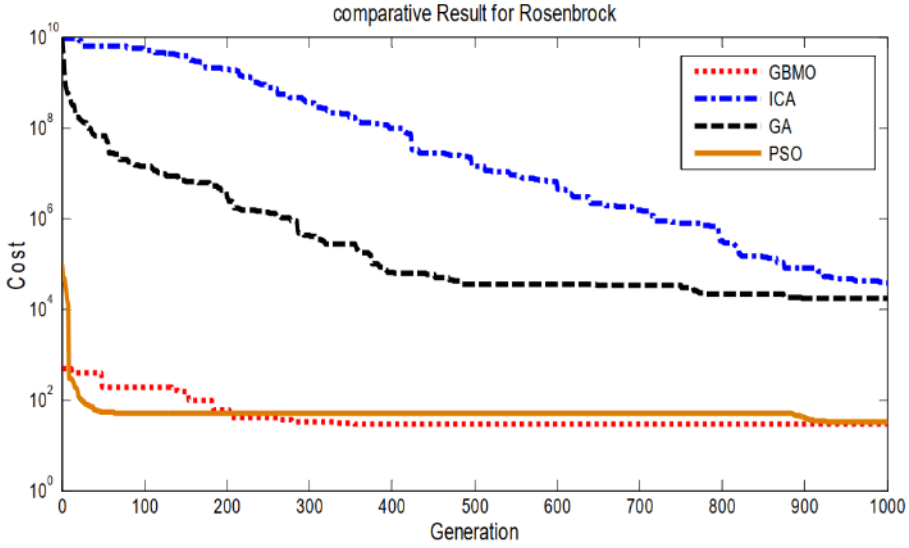

Figure 8. Results Comparison in the Rosenbrock function

But in order to evaluate the proposed approach in improving cancer detection by the combination of GBMO and KNN algorithm, three standard benchmark data have been extracted in the breast cancer field from the UCI database, where details of these data are presented in Table 6. Also, in order to compare the performance of the GBMO + KNN algorithm, three other combinations such as GA + KNN, PSO + KNN and ICA+ KNN have been tested on three breast cancer benchmark data. These algorithms are performed 10 times on the data of table 6 and the best detection accuracy, as well as the accuracy average of obtained cancer detection, is demonstrated in Tables 7 and 8, respectively.

Table 6. Utilized Breast Cancer Benchmark Data

Data	Number of Patients	Number of Symptoms	Number of Classes
WOBC	699	10	2
WDBC	569	30	2
Breast Cancer Coimbra	116	10	2

Table 7. The obtained results average for cancer detection

Data		GBMO	PSO	GA	ICA
WOBC	Test	97.79	96.15	90.54	95.63
	Train	98.41	97.18	93.15	97.44
WDBC	Test	99.34	94.16	92.10	95.21
	Train	99.59	97.81	95.46	98.07
Breast Cancer Coimbra	Test	98.15	96.30	92.22	96.26
	Train	98.57	97.05	93.77	96.94

Table 8. The obtained Best results for cancer detection

Data		GBMO	PSO	GA	ICA
WOBC	Test	99.25	98.86	97.21	98.64
	Train	99.54	99.22	97.93	98.96
WDBC	Test	99.43	98.16	95.31	97.66
	Train	99.55	97.99	97.21	98.61
Breast Cancer Coimbra	Test	98.25	98.08	96.15	97.01
	Train	98.62	98.79	97.23	97.1

The proximity of obtained results from the training and testing phase as well as the better performance of the GBMO and PSO are the most prominent achievement of evaluation and comparing the results in Tables 7 and 8.

5. Conclusion

In the past few decades, several models have been developed to identify diseases for better prediction of diseases. The purpose of the disease detection models is the decisive and premature diagnosis of the results obtained from the patient's experiments. Disease identification from medical data can help diagnose the diseases. In this paper, by focusing on the KNN classification algorithm, two challenges of this classification method were attempted to be solved by the Gases

Brownian Motion Optimization algorithm. For this purpose, the optimal selection of the appropriate subset of the training data features and k value of the nearest neighbor is obtained by the GBMO algorithm.

The results of the GBMO algorithm implementation and the other three well-known algorithms demonstrated that GBMO performs better in optimization problems with large dimensions. Furthermore, the implementation of the combination of GBMO+KNN, PSO+KNN, GA+KNN and ICA+KNN illustrated that firstly, the gap between the results of the training and testing data is very small in the implemented hybrid approaches and secondly, the GBMO and PSO algorithms have succeeded in achieving better results than other algorithms.

REFERENCES

[1] Araújo, T., et al., *Classification of breast cancer histology images using convolutional neural networks.* PloS one, 2017. **12**(6): p. e0177544.

[2] Devi, R.D.H. and M.I. Devi, *Outlier detection algorithm combined with decision tree classifier for early diagnosis of breast cancer.* Int J Adv Engg Tech/Vol. VII/Issue II/April-June, 2016. **93**: p. 98.

[3] Luo, Y., et al., *Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis.* Radiotherapy and Oncology, 2017. **123**(1): p. 85-92.

[4] Karabatak, M., *A new classifier for breast cancer detection based on Naïve Bayesian.* Measurement, 2015. **72**: p. 32-36.

[5] Dikaios, N., et al., *Logistic regression model for diagnosis of transition zone prostate cancer on multi-parametric MRI.* European radiology, 2015. **25**(2): p. 523-532.

[6] Wang, H., et al., *A support vector machine-based ensemble algorithm for breast cancer diagnosis.* European Journal of Operational Research, 2018. **267**(2): p. 687-699.

[7] Jalayeri, S. and M. Abdolrazzagh-Nezhad, *Chemical reaction optimization to disease diagnosis by optimizing hyper-planes classifiers.* Soft Computing, 2019: p. 1-20.

[8] Hossain, E., M.F. Hossain, and M.A. Rahaman. *An Approach for the Detection and Classification of Tumor Cells from Bone MRI Using Wavelet Transform and KNN Classifier*. in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. 2018. IEEE.

[9] Hiesh, M.-H., et al. *Classification of schizophrenia using genetic algorithm-support vector machine (ga-svm)*. in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. 2013. IEEE.

[10] Majid, A. and R. Homa, *Classification application with the help of cultural algorithm in predicting and diagnosing cancer.* 3rd International Conference on Applied Research in Computer Engineering and Information Technology, 2014.

[11] Wu, M., Z. Xu, and J. Watada, *Memetic algorithm based support vector machine classification.* International Journal of Innovative Management Information &Production, 2012. **3**(3): p. 99-117.

[12] Wang, K.-J., et al., *A hybrid classifier combining Borderline-SMOTE with AIRS algorithm for estimating brain metastasis from lung cancer: A case study in Taiwan.* Computer methods and programs in biomedicine, 2015. **119**(2): p. 63-76.

[13] Polat, K. and S. Güneş, *An improved approach to medical data sets classification: artificial immune recognition system with fuzzy resource allocation mechanism.* Expert Systems, 2007. **24**(4): p. 252-270.

[14] ling Chen, H., et al., *Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy.* Applied Mathematics and Computation, 2014. **239**: p. 180-197.

[15] Saidi, M., M.A. Chikh, and N. Settouti. *Automatic identification of diabetes diseases using a modified artificial immune recognition system2 (MAIRS2)*. in *Proceedings of 3ème conference internationale sur l 'informatique et ses applications*. 2011.

[16] Huang, C.-L. and J.-F. Dun, *A distributed PSO–SVM hybrid system with feature selection and parameter optimization.* Applied soft computing, 2008. **8**(4): p. 1381-1391.

[17] Ch, S., et al., *A support vector machine-firefly algorithm based forecasting model to determine malaria transmission.* Neurocomputing, 2014. **129**: p. 279-288.

[18] Ch, S., et al., *A support vector machine-firefly algorithm based forecasting model to determine malaria transmission.* Neurocomputing, 2014. **129**: p. 279-288.

[19] AlMuhaideb, S. and M.E.B. Menai, *HColonies: a new hybrid metaheuristic for medical data classification.* Applied intelligence, 2014. **41**(1): p. 282-298.

[20] Tomar, D. and S. Agarwal, *A survey on Data Mining approaches for Healthcare.* International Journal of Bio-Science and Bio-Technology, 2013. **5**(5): p. 241-266.

[21] Zheng, B., S.W. Yoon, and S.S. Lam, *Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms.* Expert Systems with Applications, 2014. **41**(4): p. 1476-1482.

[22] Han, J., J. Pei, and M. Kamber, *Data mining: concepts and techniques*. 2011: Elsevier.

[23] Jain, A.K., R.P.W. Duin, and J. Mao, *Statistical pattern recognition: A review.* IEEE Transactions on pattern analysis and machine intelligence, 2000. **22**(1): p. 4-37.

[24] Abdechiri, M., M.R. Meybodi, and H. Bahrami, *Gases Brownian motion optimization: an algorithm for optimization (GBMO).* Applied Soft Computing, 2013. **13**(5): p. 2932-2946.