Applied Soft Computing Journal 89 (2020) 106092

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Quantum based Whale Optimization Algorithm for wrapper feature selection

R.K. Agrawal [a], Baljeet Kaur [a,b,*], Surbhi Sharma [a]

[a] School of Computer and Systems Sciences, Jawaharlal Nehru University, Delhi, 110067, India
[b] Hansraj College, University of Delhi, Delhi, 110007, India

ARTICLE INFO

Article history:
Received 3 July 2019
Received in revised form 5 December 2019
Accepted 7 January 2020
Available online 16 January 2020

Keywords:
Quantum
Whale Optimization Algorithm
Bio-inspired technique
Evolutionary techniques
Swarm based techniques
Feature selection

ABSTRACT

In this paper, we propose the Quantum Whale Optimization Algorithm (QWOA) for feature selection, which is an amalgamation of the Quantum Concepts and the Whale Optimization Algorithm (WOA). The proposed method enhances the exploratory and exploitation power of the classical WOA, with the use of quantum bit representation of the individuals of the population and the quantum rotation gate operator as a variation operator. Modified mutation and crossover operators are also introduced for quantum-based exploration, shrinking and spiral movement of the whales in the proposed QWOA. The efficacy of the proposed method is compared with that of the conventional WOA and with well-known evolutionary, swarm and quantum algorithms with fourteen datasets from diversified domains. Experimental results demonstrate the superior performance of the proposed QWOA method. Statistical tests also demonstrate the significantly better performance of the QWOA in comparison to eight well-known meta-heuristic algorithms.

1. Introduction

All natural processes are inherently optimal in behaviour, since they drive an organism and/ or a complete social setup towards survival. The principle of natural selection and evolution; and the strategies that various living organisms pursue to search for food have become the basis of many evolutionary algorithms and swarm computing techniques.

The evolutionary algorithms are characterized by (a) proper representation of the individuals within a population, (b) the fitness function that evaluates the goodness of the individual, (c) variation operators like mutation and crossover that update the individuals and drive them towards optimal solution and (d) other parameters such as the size of the population, the selection strategy etc. The judicious search space exploration and exploitation of the probable candidate solutions are instrumental in the timely convergence of the evolutionary algorithms. Genetic Algorithm (GA) [1] and Genetic Programming [2] are popular evolutionary algorithms, which have successfully been applied to the areas of signal processing [3], wireless communication [4], and financial modelling [5] to name a few.

Nature also inspires swarm based computing techniques that are motivated by the social behaviour of animals that work in a team to forage for food. Swarm-based techniques like Ant Colony Optimization (ACO) [6], Particle Swarm Optimization (PSO) [7] and Artificial Bee Colony Optimization (ABC) [8] are based upon the social behaviour of living organisms. The Bat algorithm (BA) [9], the Grey Wolf Optimization Algorithm (GWO) [10] and the Swarm Salp Algorithm (SSA) [11] are some recently proposed swarm algorithms. They too propel the system towards an optimal solution, through repeated exploration and exploitation techniques. Various applications of swarm computation include aircraft scheduling [12], crowd simulation [13], medical imaging [14] among others. Biological evolution and natural adaptation have laid down the foundation of evolutionary algorithms which have seen phenomenal progress in their role in optimization of functions [15–19] and feature subset search techniques using GA [20,21], PSO [22,23], ACO [24], Binary Bat Algorithm (BBA) [9], GWO [25] and SSA [11].

The application of feature selection plays a pivotal role in building efficient and improved decision systems. Feature selection techniques help to obtain the relevant and non-redundant features, which reduce the complexity of the system in terms of time and space. The reduced relevant features also provide better performance. Methods for feature selection are broadly categorized into filter and wrapper [26,27]. The filter feature selection methods are based on the statistical properties of the individual features. As these are independent of the classifier, they are scalable but do not guarantee to perform well with a given

* Corresponding author at: Hansraj College, University of Delhi, Delhi, 110007, India.
E-mail addresses: rkajnu@gmail.com (R.K. Agrawal), baljeetkaur26@hotmail.com (B. Kaur), surbhisharma9099@gmail.com (S. Sharma).

https://doi.org/10.1016/j.asoc.2020.106092
1568-4946/

learning algorithm. Wrapper methods select a subset of features based on the evaluation criteria of a learning algorithm. Further, wrapper methods may be sequential or random in nature [27]. In sequential wrapper methods, like the forward feature selection and backward elimination method, features are included and removed respectively in the selected subset, incrementally. However, the features so selected/removed are not removed/selected in further iterations to enhance the performance of the decision system by these methods. Random search based wrapper methods overcome the limitations of the sequential wrapper methods as they provide an improved strategy for exploring the feature space. Some well known random search based wrapper methods are the Genetic Algorithm [1], Ant Colony Optimization [6] and Particle Swarm Optimization [7] and. These algorithms have been studied due to their metaheuristic capabilities. In these methods, a population of probable solutions is generated, which converges towards the optimal solution, over generations.

Recently, a feature selection approach [28] has been proposed, based on the Whale Optimization Algorithm (WOA) [29]. The WOA is based on the principles of the mathematical formulation of the prey searching strategy of the humpback whales. The WOA explores the domain space through a random search, as well as a guided search, based on the whale's technique of encircling and closing in on the prey. Feature selection based WOA [28] involves the mutation and crossover operators, for enhancing exploitation, along with the Tournament selection, for a thorough exploration. Results on benchmark datasets show improvements over other algorithms like the GA, PSO and ALO.

Recently, there is an exponential surge in the dimension of the feature space, which encourages the application of feature selection algorithms that are time and space efficient. However, the classical evolutionary and swarm-based feature selection methods involve huge computation overheads due to the large population size and the representation of the individuals of the population.

To handle the time and space complexity, the principle of quantum computing is amalgamated with some nature-inspired techniques in the recent past [30–34]. Quantum computing is based on the principles of the quantum bit [30] and promotes population diversity due to the superposition of states while facilitating the convergence within a short span of time. Quantum computing is successfully combined with the Genetic Algorithm [31] to experience parallel computing. The Quantum Evolutionary Algorithm (QEA) for combinatorial optimization [30] outperformed the conventional GA, without premature convergence even when small population size was used. Sun et al. have proposed a quantum version of PSO [34] and established its efficacy using some benchmark functions. Quantum-inspired GA (QGA) [32], as well as the Quantum inspired PSO (QPSO) [34] have succeeded as the preferred feature selection techniques over their conventional counterparts.

Motivated by the research works [30–34], we propose to enhance the performance of the classical Whale Optimization Algorithm with the inclusion of the quantum computing principle. In the proposed Quantum Whale Optimization Algorithm (QWOA) method, the Q-bit (quantum bit) representation, the quantum rotation gate, and the modified evolutionary operators like selection, mutation, and crossover are utilized. The performance of the QWOA is compared with the classical WOA and seven well-known meta-heuristic methods such as GA, PSO, BBA, GWO, SSA, QGA and QPSO in terms of a fitness function (based on the classification error and number of selected features), classification accuracy, area under the ROC curve and the selected feature ratio on fourteen challenging datasets from diversified domains. These datasets have been chosen from different domains of microarray gene expression, facial images, text and a variety of UCI datasets. The feature dimension of these datasets ranges from tens up to many thousands. For the high dimensional datasets, the QWOA and other meta-heuristic methods incur high time and space cost. To handle the complexity involved with high dimension datasets, prior to the implementation of the meta-heuristic method, a clustering method is employed that chooses a relevant and non-redundant feature set to be input to the method. This two-stage implementation improves the performance of the methods with high dimensional data. We have used four well-known classifiers such as the *k*-Nearest Neighbour (kNN), the Linear discriminant classifier (LDC), the Support Vector Machine (SVM) and the C4.5. A statistical significance test is also performed to evaluate the statistical significance of the proposed method in comparison to the eight well-known evolutionary, swarm and quantum algorithms.

Following are the main contributions of our work:

1. A quantum-based approach of the classical WOA, QWOA, is proposed that further enhances the diversification and convergence properties of this swarm based wrapper feature selection method, through the quantum-bit representation of the population.
2. Modified mutation and crossover operators are introduced for quantum-based exploration, shrinking and spiral movement of the whales.
3. The inclusion of the quantum rotation gate operator boosts the convergence to a single state, hence balancing the exploration (diversification) and exploitation (convergence) property of the proposed method.
4. Reduction of the feature input set using the clustering step prior to QWOA for high dimensional datasets is instrumental in building a high performing decision system.
5. Performance comparison of the proposed method with eight popular evolutionary, swarm and quantum algorithms on fourteen datasets is carried out.

A brief introduction of the Whale Optimization Algorithm and the characteristics of the proposed Quantum Whale Optimization Algorithm (QWOA) for improved feature selection are presented in Section 2. Experimental setup, results, and discussions are presented in Section 3. The conclusion section and future work conclude the paper.

2. Proposed QWOA for feature selection

2.1. Whale optimization algorithm for feature selection

Whale Optimization Algorithm [29] is a recently proposed nature inspired optimization method that emulates the hunting behaviour of the humpback whales. Humpback whales are characterized by the exploration, shrinking and spiralling behaviour while hunting, which has been mathematically modelled in the WOA.

Consider $\boldsymbol{W}$ to be a randomly generated set of solutions (whale positions) having n whales in a d-dimensional search space. While implementing WOA for feature selection [28], each whale position corresponds to a feature vector of length d, where value 1 or 0 of a feature indicates its inclusion or exclusion respectively in the feature subset. In each generation, the performance of each of the whale position, $\boldsymbol{W}^i$ (i=1, 2, …, n) is assessed based on some fitness criteria. After calculating the fitness of each solution, $\boldsymbol{W}^i$, the fittest whale position, $\boldsymbol{W}^*$ is ascertained. The positions of the n whales are updated in every generation till a terminating condition is reached. The updation may be in the form of exploration, shrinking or spiralling.

In *exploration*, a random candidate, $\boldsymbol{W}^r$ is chosen from the population to guide the search. The current whale, $\boldsymbol{W}^i$, is mutated and crossed over with the randomly chosen whale position to

Algorithm : Quantum Whale Optimization Algorithm for Feature Selection

t ← 1
T ← *maxGen*
Initialize a feature population of size n,
$\mathbf{W}(t)=\{\mathbf{W}^1(t), \mathbf{W}^2(t), \ldots, \mathbf{W}^i(t), \ldots, \mathbf{W}^n(t)\}^T$
where $\mathbf{W}^i(t)=\{ W^{i1}(t), W^{i2}(t), \ldots, W^{ij}(t), \ldots, W^{id}(t)\}$, d: dim. of quantum individual
$$W^{ij}=\begin{bmatrix} a_{ij} \\ b_{ij} \end{bmatrix} is\ a\ Q-bit,\ s.t.\ |a_{ij}|^2 + |b_{ij}|^2 = 1$$
Observe/Collapse$\mathbf{W}(t)$ to give the binary population $\mathbf{B}(t)$ (Equation 6)
Calculate Fitness value for each binary feature vector$\mathbf{B}^i(t)$, *i=1,2,3 ... n*
$\mathbf{B}^*(t)$←BestFitnessVector($\mathbf{B}(t)$)
bestIndex= BestFitnessIndex($\mathbf{B}(t)$)
$\mathbf{W}^*(t)$←$\mathbf{W}^{bestIndex}(t)$
While t< T
  For each individual, *i = 1,2,3, ... n*
    Choose a random value p and update *A*(Equation 4)
    If p < threshold
      If abs(A)≥ 1 ***//Explore***
        $\mathbf{W}^r(t)$←TouramentSelection($\mathbf{W}(t)$)
        $\mathbf{W}^r(t)$←Mutation($\mathbf{W}^r(t)$)
        $\mathbf{W}^i(t)$←Crossover(Mutation($\mathbf{W}^i(t)$),$\mathbf{W}^r(t)$)
      End If
      If abs(A) <1 ***//Shrink***
        $\mathbf{W}^i(t)$←Crossover(Mutation($\mathbf{W}^*(t)$), $\mathbf{W}^*(t)$)
      End If
    End If
    If p ≥threshold ***//Spiral***
      $\mathbf{D}$←distance ($\mathbf{W}^i(t)$, $\mathbf{W}^*(t)$)
      $\mathbf{W}^i(t)$←Update i^{th} individual (Equation 2)
    End If
End For
  Update $\mathbf{W}(t)$ using rotation gate (Equation 10)
  Observe/Collapse$\mathbf{W}(t)$ to give the binary population $\mathbf{B}(t)$ (Equation 6)
  Calculate Fitness value for each binary feature vector$\mathbf{B}^i(t)$, *i= 1,2, ... n*
  $\mathbf{B}^*(t)$←BestFitnessVector($\mathbf{B}(t)$)
  bestIndex= BestFitnessIndex($\mathbf{B}(t)$)
  $\mathbf{W}^*(t)$←$\mathbf{W}^{bestIndex}(t)$
  t←t+1
End While
Return optimal Feature Vector $\mathbf{B}^*(t-1)$

Fig. 1(a). QWOA algorithm.

give the updated position of the current whale. While in *shrinking*, the best whale position, $\boldsymbol{W}^*$ is mutated and crossed-over with the current whale, $\boldsymbol{W}^i$,to give the updated position of the current whale. For the *spiral* movement, the distance vector between the best whale position and the current whale is calculated as:

$$\mathbf{D} = \boldsymbol{W}^* - \boldsymbol{W}^i \tag{1}$$

The updated position on spiralling is computed as:

$$\boldsymbol{W}^i = \mathbf{D} \cdot e^{pl} \cdot cos\,(2\pi l) + \boldsymbol{W}^* \tag{2}$$

where p defines the shape of the logarithmic spiral. The value of l is chosen in the interval [-1, 1]. The dot operator represents the element by element multiplication.

In each generation, the decision as to which of the three options of exploration, spiralling or shrinking, is to be performed, is based on a random choice, $p \in [0,1]$ and the value of the coefficient $\boldsymbol{A}$. To calculate $\boldsymbol{A}$, a parameter q is considered, which decreases linearly from 2 to 0 over the generations, using the following:

$$q = 2 - t * 2/maxGen \tag{3}$$

where t is the current generation and *maxGen* is the maximum generations for which the method is run. Coefficient vector $\boldsymbol{A}$ is defined to control the search space around the best solution. It is given as:

$$\boldsymbol{A} = 2q \cdot \boldsymbol{r} - q \tag{4}$$

where $\boldsymbol{r}$ is a random vector.

The classical WOA for feature selection briefly introduced above, has shown improvement over other evolutionary algorithms like PSO, ALO and GA [28], for some problems. The WOA leads the search either by using the best solution obtained or, by choosing a random whale position. This facilitates in the exploratory behaviour of WOA and prevents the system from getting stuck in the local minima.

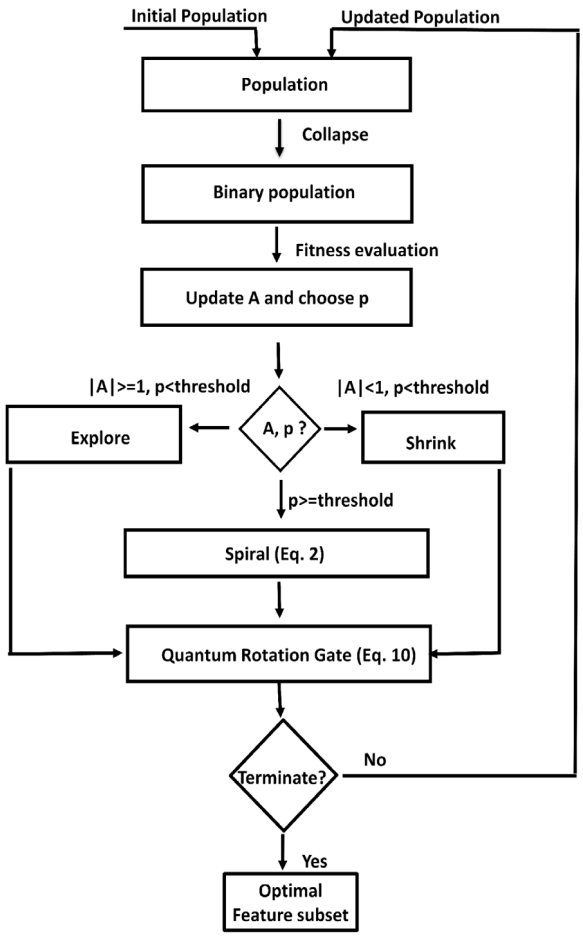

Fig. 1(b). Flowchart of the proposed QWOA method for feature selection.

Feature selection based on WOA uses the binary representation of the feature vectors. Hence the population size required for an effective search is high, which leads to more exploratory and exploitation time to converge to an optimal solution. To handle the complexity of the classical WOA in terms of time and space, to select an optimal number of features, we propose the quantum-based WOA (QWOA) which is based on the quantum computing principles and takes the advantage of a probabilistic representation of the Q-bits and enhances the population diversity. The proposed quantum-based WOA (QWOA) is motivated by the promising results for feature selection achieved by the quantum-based evolutionary algorithms, like QGA [32] and QPSO [33]. The amalgamation of quantum computing and the Whale Optimization Algorithm for feature selection is being attempted for the first time. The proposed principle of the QWOA is presented next, which involves the Q-bit representation, quantum rotation operator and modified mutation and crossover operators.

2.2. QWOA for feature selection

Quantum computing is based on the representation of data as quantum bits (Q-bit) and capitalizes on the advantage offered by the superposition of states. The Q-bit is the basic information unit of a 2-state quantum computer [31]. Given individual states, $|0\rangle$ and $|1\rangle$, the quantum state $|\Psi\rangle = a|0\rangle + b|1\rangle$ is the linear superposition of the individual states. The probability amplitudes, a and b are the complex number weights of the quantum particle in location $|0\rangle$ and $|1\rangle$ respectively. Here, $|a|^2$ and $|b|^2$ give the probability of the Q-bit to be in the state $|0\rangle$ and $|1\rangle$ respectively such that:

$$|a|^2 + |b|^2 = 1 \tag{5}$$

If a representation is composed of d states, then this representation is said to be at all of the 2^d states at the same time, each with a given probability whose sum is equal to 1. The d-length Q-bit individual can be collapsed to a d-length binary vector using the following [30]:

$$if\ |a_i|^2 < threshold\ then\ y_i \leftarrow 1\ else\ y_i \leftarrow 0 \tag{6}$$

where y_i represents the corresponding binary bit of the observed Q-bit. The *threshold* value maybe generated randomly. The set of binary vectors obtained are evaluated for fitness.

Variation operators such as selection, mutation, crossover and quantum gates are used to update the position of the Q-bits in subsequent generations. Selection operators like Best, Random, Roulette, and Tournament [35] are used to select those individuals that participate in the generation of the next set of population. The mutation operator involves altering the probabilities of one or more number of Q-bits of an individual, as explained below. For example, for a quantum individual $\boldsymbol{P}$, the mutated individual, $\boldsymbol{P}_{mut}$ at the mutation point (in bold) is represented as follows:

$$\boldsymbol{P} = \begin{pmatrix} a_1\, a_2\, \boldsymbol{a_3}\, a_4\, a_5 \ldots a_d \\ b_1\, b_2\, \boldsymbol{b_3}\, b_4\, b_5 \ldots b_d \end{pmatrix} \boldsymbol{P}_{mut} = \begin{pmatrix} a_1\, a_2\, \boldsymbol{a_3}\, a_4\, a_5 \ldots a_d \\ b_1\, b_2\, \boldsymbol{b_3}\, b_4\, b_5 \ldots b_d \end{pmatrix} \tag{7}$$

Crossover operates on two quantum individuals and outputs two individuals based on either 1-point, n-point crossover or

Table 1
Datasets used and their details.

Dimension	Domain	Dataset	Class	Samples	Original features	Pre-processed features (m)	Features after clustering
High	Microarray	GLI-85	2	85	22283	5000	100
High	Microarray	GLA_BRA180	4	180	49151	5000	100
High	Microarray	GCM	14	198	16063	5000	100
High	Microarray	Tumor-9	9	60	5726	5000	100
High	Face Image	AR10P	10	130	2400	2400	100
High	Face Image	PIE10P	10	210	2400	2400	100
High	Face Image	PIX10P	10	100	10000	2400	100
High	Text	TR11WC	9	414	6430	6430	200
High	Text	TR23WC	6	204	5833	5833	200
Low	UCI	Lymphography	4	148	18	18[b]	–
Low	UCI	Waveform	3	5000	40	40[b]	–
Low	UCI	Tictactoe	2	958	27[a]	27[b]	–
Low	UCI	StatlogHeart	2	270	13	13[b]	–
Low	UCI	Ionosphere	2	351	34	34[b]	–

[a] Encoded.
[b] All features selected.

Table 2(a)
Comparison between the proposed QWOA method and other competitive methods in terms of fitness values.

Dataset	Measure	GA	PSO	BBA	GWO	SSA	QGA	QPSO	WOA	QWOA
Lymphography	avg	0.1752	0.1802	0.1847	0.1538	0.1672	0.1469	0.1630	0.1538	**0.1445**
	std	0.0373	0.0106	0.0226	0.0163	0.0120	0.0104	0.0166	0.0110	0.0109
Waveform	avg	0.1723	0.1763	0.2001	0.1673	0.1793	0.1708	0.1707	0.1692	**0.1662**
	std	0.0036	0.0047	0.0063	0.0042	0.0022	0.0042	0.0052	0.0048	0.0032
Tictactoe	avg	0.1247	0.1135	0.1291	0.1124	0.1120	0.1102	0.1103	0.1089	**0.1085**
	std	0.0033	0.0035	0.0061	0.0033	0.0021	0.0009	0.0017	0.0004	0.0000
StatlogHeart	avg	0.1547	0.1493	0.1734	0.1441	0.1504	0.1476	0.1429	0.1422	**0.1366**
	std	0.0166	0.0085	0.0084	0.0071	0.0059	0.0048	0.0074	0.0065	0.0039
Ionosphere	avg	0.1045	0.1156	0.1141	0.1145	0.1186	0.0921	0.1039	0.0838	**0.0754**
	std	0.0232	0.0083	0.0101	0.0071	0.0041	0.0084	0.0107	0.0135	0.0092
GLI-85	avg	0.0073	0.0077	0.0197	0.0055	0.0049	0.0047	0.0064	0.0033	**0.0028**
	std	0.0078	0.0054	0.0077	0.0007	0.0005	0.0008	0.0034	0.0006	0.0007
GLA-BRA180	avg	0.2179	0.2147	0.2392	0.1977	0.2174	0.2080	0.2013	0.2080	**0.1960**
	std	0.0243	0.0121	0.0275	0.0106	0.0094	0.0128	0.0141	0.0119	0.0118
GCM	avg	0.3532	0.4108	0.4973	0.3383	0.3891	0.3488	0.3836	0.2869	**0.2830**
	std	0.0478	0.0243	0.0322	0.0303	0.0239	0.0297	0.0290	0.0447	0.0256
Tumour-9	avg	0.2703	0.2924	0.3585	0.2526	0.2975	0.2660	0.2726	0.2460	**0.2124**
	std	0.0298	0.0234	0.0334	0.0307	0.0157	0.0281	0.0186	0.0188	0.0183
AR10P	avg	0.0945	0.2138	0.2753	0.1844	0.2279	0.0883	0.2121	0.0112	**0.0074**
	std	0.0221	0.0188	0.0129	0.0216	0.0118	0.0374	0.0306	0.0064	0.0040
PIE10	avg	0.0051	0.0059	0.0152	0.0063	0.0053	0.0048	0.0058	0.0041	**0.0034**
	std	0.0015	0.0023	0.0039	0.0004	0.0004	0.0010	0.0007	0.0007	0.0007
PIX10P	avg	0.0110	0.0116	0.0120	0.0126	0.0113	0.0091	0.0106	0.0048	**0.0043**
	std	0.0013	0.0039	0.0034	0.0040	0.0046	0.0097	0.0053	0.0030	0.0049
TR11WC	avg	0.2042	0.2236	0.2559	0.2050	0.2255	0.1870	0.2151	0.1820	**0.1769**
	std	0.0055	0.0105	0.0126	0.0086	0.0052	0.0116	0.0080	0.0096	0.0099
TR23WC	avg	0.2097	0.2223	0.2457	0.2062	0.2179	0.1764	0.2005	0.1422	**0.1404**
	std	0.0086	0.0148	0.0142	0.0070	0.0063	0.0318	0.0092	0.0180	0.0181
Average		0.1503	0.1670	0.1943	0.1501	0.1660	0.1401	0.1571	0.1247	**0.1184**

Table 2(b)
Comparison between the proposed QWOA method and other competitive methods in terms of classification accuracy.

Dataset	Meas	GA	PSO	BBA	GWO	SSA	QGA	QPSO	WOA	QWOA
Lymphography	avg	82.7682	82.5000	81.8243	85.1352	83.7162	85.8085	84.1216	85.0000	**86.0135**
	std	3.8907	1.0778	2.1473	1.6551	1.1682	1.0560	1.6929	1.1832	1.1057
Waveform	avg	83.1586	83.0640	80.3600	83.8520	82.5060	83.2485	83.3340	83.6640	**83.8660**
	std	0.3274	0.4439	0.6034	0.3911	0.2565	0.3811	0.5173	0.4239	0.2742
Tictactoe	avg	87.9799	89.3215	87.5574	89.3842	89.3737	89.5111	89.5199	**89.5616**	**89.5616**
	std	0.3930	0.3191	0.6761	0.2871	0.2350	0.1619	0.1320	0.0012	0.0010
StatlogHeart	avg	84.9991	85.6667	83.1111	86.2222	85.3333	85.8455	86.2222	86.3704	**86.8519**
	std	1.7554	0.8562	0.9434	0.7366	0.5576	0.3349	0.7156	0.5466	0.2619
Ionosphere	avg	89.8872	88.8034	88.7465	88.9744	88.4331	90.8326	89.8576	91.8234	**92.9900**
	std	2.1371	0.7484	0.9708	0.6448	0.4074	0.8375	1.0068	1.2313	0.9153
GLI-85	avg	99.5290	99.6471	98.3530	**100.0000**	**100.0000**	99.9993	99.8824	**100.0000**	**100.0000**
	std	0.8231	0.5683	0.8226	0.0000	0.0000	0.0014	0.3720	0.0000	0.0000
GLA-BRA180	avg	78.3141	78.8333	76.3333	80.8206	78.6111	79.6556	80.3333	79.6919	**80.8805**
	std	2.5811	1.1843	2.7492	1.6276	0.9533	1.2773	1.4152	1.1548	1.1057
GCM	avg	64.7832	59.1305	50.2174	66.5217	61.3044	65.4337	61.7392	71.5217	**71.9565**
	std	4.8934	2.4680	3.3128	3.1083	2.4680	2.9783	2.9346	4.5195	2.6027
Tumour-9	avg	73.1465	71.0000	64.3333	75.1667	70.5000	73.8330	73.1667	75.3000	**79.1600**
	std	2.9860	2.3831	3.3518	3.0882	1.5812	2.8385	1.8342	1.8002	1.8159
AR10P	avg	90.8101	79.0769	72.6154	82.0769	77.5385	91.5990	79.0769	99.4615	**99.8462**
	std	2.1696	1.8419	1.3175	2.1772	1.1917	3.6940	3.0512	0.6333	0.3243
PIE10	avg	99.8990	99.5238	98.9524	**100.0000**	**100.0000**	**100.0000**	**100.0000**	**100.0000**	**100.0000**
	std	0.0000	0.0000	0.4376	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PIX10P	avg	99.1040	99.2000	99.1000	99.2000	99.3000	90.1697	99.4000	**99.9000**	99.7000
	std	0.1384	0.4216	0.3162	0.4216	0.4830	28.1853	0.5164	0.3162	0.4830
TR11WC	avg	79.8838	77.9227	74.6377	80.0725	77.8261	82.0000	78.9855	82.4155	**82.8744**
	std	0.5483	1.0694	1.2933	0.8765	0.5437	1.1297	0.7467	0.9095	0.9625
TR23WC	avg	79.2409	78.2843	75.5392	79.9020	78.5294	82.8177	80.2451	86.2745	**86.3725**
	std	0.9157	1.4440	1.4514	0.6932	0.6454	3.2002	0.9258	1.7750	1.8019
Average		85.2503	83.7124	80.8344	85.5234	83.7837	85.7682	84.7060	87.9275	**88.5766**

Table 2(c)
Comparison between the proposed QWOA method and other competitive methods in terms of AUC.

Dataset	Measure	GA	PSO	BBA	GWO	SSA	QGA	QPSO	WOA	QWOA
Lymphography	avg	0.8031	0.8204	0.6924	0.8349	0.8314	0.8438	0.8312	0.8282	**0.8520**
	std	0.0710	0.0277	0.0817	0.0302	0.0335	0.0127	0.0109	0.0210	0.0209
Waveform	avg	0.9116	0.9462	0.7904	0.9494	0.9460	0.9205	0.9483	0.9496	**0.9511**
	std	0.0020	0.0021	0.1414	0.0029	0.0028	0.0026	0.0030	0.0018	0.0024
Tictactoe	avg	0.9697	0.9692	0.7695	0.9712	0.9728	0.9712	0.9729	0.9732	**0.9737**
	std	0.0133	0.0040	0.0898	0.0037	0.0026	0.0018	0.0027	0.0025	0.0022
StatlogHeart	avg	0.8670	0.8713	0.8006	0.8823	0.8751	0.8694	0.8813	0.8763	**0.8838**
	std	0.0264	0.0163	0.0477	0.0107	0.0160	0.0084	0.0141	0.0156	0.0179
Ionosphere	avg	0.9214	0.9182	0.9103	0.9244	0.9241	0.9249	0.9264	0.9343	**0.9349**
	std	0.0156	0.0077	0.0104	0.0144	0.0118	0.0091	0.0153	0.0141	0.0149
GLI-85	avg	0.9931	0.9958	0.9553	0.9993	0.8878	0.9950	0.9993	0.9996	**0.9998**
	std	0.0156	0.0052	0.0609	0.0015	0.0162	0.0143	0.0015	0.0013	0.0012
GLA-BRA180	avg	0.8819	0.8861	0.7970	0.8935	0.8092	0.8908	0.8953	0.8998	**0.9010**
	std	0.0249	0.0146	0.0742	0.0128	0.0272	0.0099	0.0093	0.0126	0.0006
GCM	avg	0.8160	0.7996	0.5527	0.8084	0.8092	0.8282	0.7959	0.8259	**0.8322**
	std	0.0271	0.0261	0.2256	0.0229	0.0272	0.0295	0.0244	0.0300	0.0203
Tumour-9	avg	0.7323	0.7408	0.6757	0.7473	0.7412	0.7573	0.7412	0.7357	**0.7698**
	std	0.0154	0.0095	0.0644	0.0113	0.0159	0.0124	0.0148	0.0151	0.0123
AR10P	avg	0.9696	0.9331	0.8506	0.9444	0.9411	0.9704	0.9502	0.9978	**0.9982**
	std	0.0104	0.0083	0.1046	0.0116	0.0103	0.0119	0.0151	0.0025	0.0025
PIE10	avg	0.9998	0.9997	0.9654	0.9999	0.9998	0.9999	0.9998	**0.9998**	**0.9998**
	std	0.0002	0.0003	0.0615	0.0001	0.0002	0.0001	0.0001	0.0002	0.0002
PIX10P	avg	0.9970	0.9959	0.9867	0.9958	0.9969	0.9993	0.9969	**0.9998**	0.9990
	std	0.0035	0.0034	0.0135	0.0033	0.0033	0.0012	0.0035	0.0006	0.0026
TR11WC	avg	0.9176	0.9075	0.7731	0.9182	0.9128	0.9212	0.9209	0.9207	**0.9322**
	std	0.0158	0.0141	0.0950	0.0100	0.0143	0.0097	0.0061	0.0083	0.0053
TR23WC	avg	0.8884	0.8629	0.7743	0.8667	0.8659	0.8909	0.8750	0.8866	**0.8917**
	std	0.0199	0.0093	0.0678	0.0198	0.0192	0.0206	0.0259	0.0183	0.0204
Average		0.9049	0.9033	0.8067	0.9097	0.8938	0.9131	0.9096	0.9162	**0.9228**

Table 2(d)
Comparison between the proposed QWOA method and other competitive methods in terms of the minimum number of selected features.

Dataset	Measure	GA	PSO	BBA	GWO	SSA	QGA	QPSO	WOA	QWOA
Lymphography	avg	15.2	12.6	**8.6**	11.9	10.8	**8.6**	10.5	9.6	10.9
	std	10.71	1.84	3.31	1.1	2.25	0.84	1.72	2.72	1.45
Waveform	avg	29.2	34.7	32.5	29.8	24.4	**17**	22.9	29.9	25.9
	std	12.63	1.95	2.84	1.75	3.13	1.62	2.13	3.51	2.69
Tictactoe	avg	22.5	21	15.9	19.7	18.4	14.8	17.7	15	**14**
	std	13.11	1.56	2.38	1.83	1.35	1.55	1.83	1.15	0
StatlogHeart	avg	15.1	9.6	8	10	**6.7**	7.2	8.5	9.4	8.4
	std	10.11	1.65	2.87	1.05	1.16	1.55	2.51	1.78	2.32
Ionosphere	avg	21.8	16.1	9	18.2	14	**1.6**	12	9.6	6.1
	std	14.3	3.75	2.94	2.97	3.16	0.7	3.62	4.7	1.2
GLI-85	avg	23	31.4	32.1	33.2	29.6	25.4	25.1	19.8	**17**
	std	14.9	7.07	4.89	4.16	2.95	1.99	6.79	3.52	4.08
GLA-BRA180	avg	**26.5**	39.6	29.3	43.6	33.8	36.3	34.8	41.7	40.3
	std	17.5	4.27	3.92	2.07	3.91	4.12	3.26	7.96	6.55
GCM	avg	34.5	37.4	47	41.4	36	36.8	32.1	32.3	**30**
	std	16.94	7.37	6.45	3.17	6.65	0.91	4.58	3.02	4.14
Tumour-9	avg	53.7	41.7	**32.5**	40.6	32.7	38.5	35.5	43.3	37.1
	std	19.28	5.48	3.95	3.78	2.5	3.34	3.37	7.2	2.94
AR10P	avg	52.5	66.2	51.4	69.5	55.1	**44.5**	49.4	59	58.3
	std	24.98	7.36	9.87	4.38	2.51	14.38	4.84	5.87	10.04
PIE10	avg	48.1	57.9	58.3	63.4	53.1	54.9	44.6	40.6	**36.2**
	std	22.47	6.71	7.45	4.25	3.96	6.99	3.66	6.88	7.27
PIX10P	avg	28.3	46.8	41	46.5	43.8	**8**	36.7	18.3	12.8
	std	14.96	6.39	12.43	5.28	4.1	9.72	5.92	3.06	3.39
TR11WC	avg	107.6	141.5	**96.3**	154.4	118.8	144.4	100.1	158.3	122.2
	std	60.62	23.13	16.81	8.45	6.99	5.77	5.88	19.2	29.1
TR23WC	avg	90.1	146.7	**70.3**	143.9	107.1	100.7	105.1	126.8	110.7
	std	51.74	15.62	18.37	7.78	7.92	23.6	6.3	17.29	22.4
Average		40.58	50.229	38.014	51.86	41.736	38.479	38.214	43.829	**37.85**

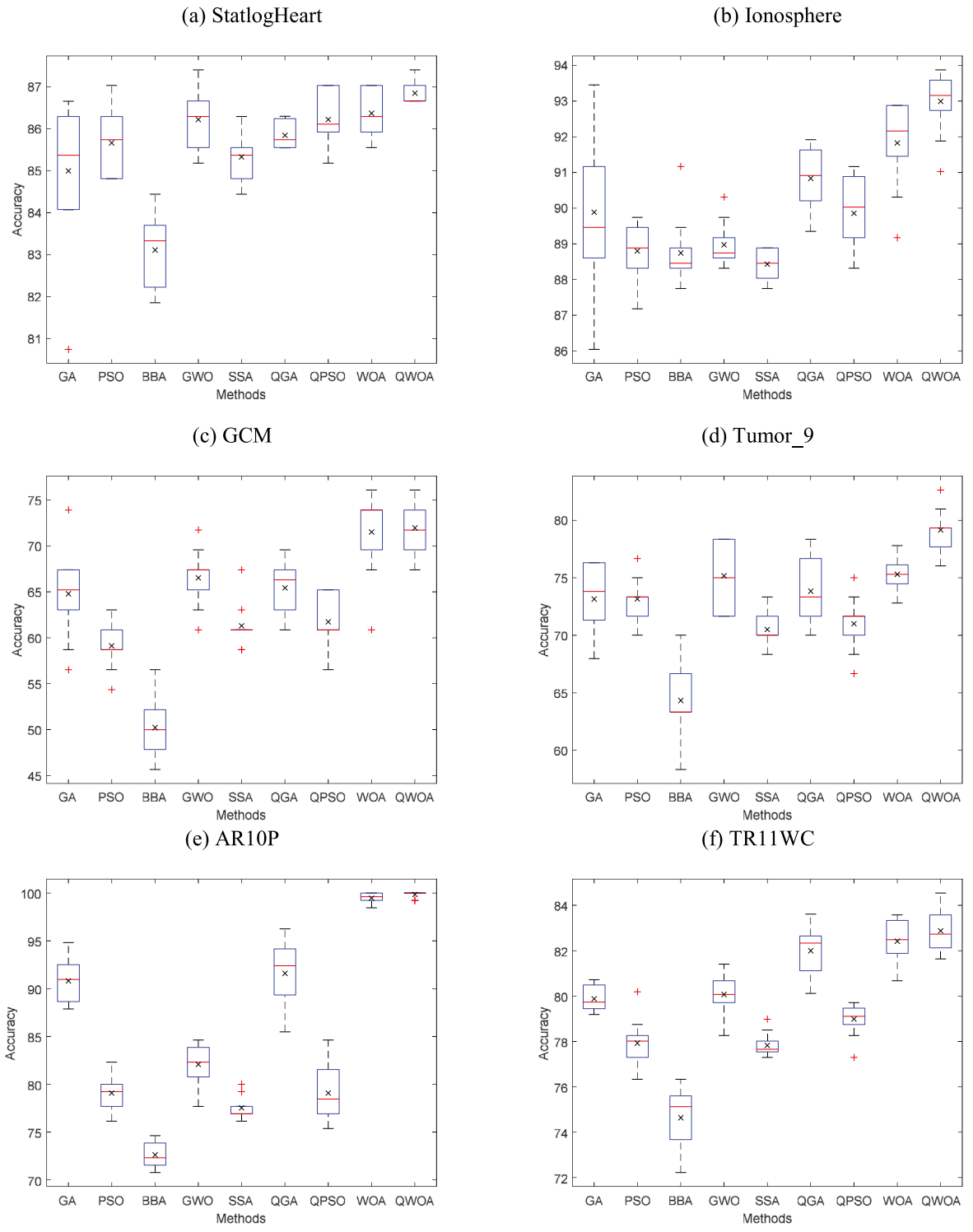

Fig. 2. Boxplots of classification accuracy for QWOA versus other methods for (a) StatlogHeart (b) Ionosphere (c) GCM (d) Tumor_9 (e) AR10P and (f) TR11WC.

random point crossover. Given below is an illustration for a one-point crossover at the third position between parents $\boldsymbol{P}^1$ and $\boldsymbol{P}^2$ to give children $\boldsymbol{C}^1$ and $\boldsymbol{C}^2$.

$$\boldsymbol{P}^1 = \begin{pmatrix} a_1^1\, a_2^1\, \boldsymbol{a_3^1}\, a_4^1\, a_5^1 \ldots a_{\mathrm{d}}^1 \\ b_1^1\, b_2^1\, \boldsymbol{b_3^1}\, b_4^1\, b_5^1 \ldots b_{\mathrm{d}}^1 \end{pmatrix} \boldsymbol{P}^2 = \begin{pmatrix} a_1^2\, a_2^2\, \boldsymbol{a_3^2}\, a_4^2\, a_5^2 \ldots a_{\mathrm{d}}^2 \\ b_1^1\, b_2^2\, \boldsymbol{b_3^2}\, b_4^2\, b_5^2 \ldots b_{\mathrm{d}}^2 \end{pmatrix} \tag{8}$$

$$\boldsymbol{C}^1 = \begin{pmatrix} a_1^1\, a_2^1\, a_3^2\, a_4^2\, a_5^2 \ldots a_{\mathrm{d}}^2 \\ b_1^1\, b_2^1\, b_3^2\, b_4^2\, b_5^2 \ldots b_{\mathrm{d}}^2 \end{pmatrix} \boldsymbol{C}^2 = \begin{pmatrix} a_1^2\, a_2^2\, a_3^1\, a_4^1\, a_5^1 \ldots a_{\mathrm{d}}^1 \\ b_1^1\, b_2^2\, b_3^1\, b_4^1\, b_5^1 \ldots b_{\mathrm{d}}^1 \end{pmatrix} \tag{9}$$

The application of a quantum gate operator [30] on the probability amplitude of each Q-bit individual balances exploitation and exploration of the QWOA to speed up the search and giving the result within a shorter span of time. Commonly used quantum gates are the Hadamard gate, the NOT gate, the controlled-NOT gate, and the rotation gate [30]. We have used the rotation quantum gate:

$$\boldsymbol{U}(\varphi) = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \tag{10}$$

Table 3
Variation in performance based on population size.

Dataset	Pop.Size	Average fitness		Average accuracy		Average features	
		WOA	QWOA	WOA	QWOA	WOA	QWOA
GLA_BRA180	1	0.262	**0.251**	74.17	**75.33**	39.3	**37.8**
	10	0.218	**0.216**	78.67	**78.89**	42.1	**40.3**
	20	0.208	**0.196**	79.67	**80.89**	41.7	**40.3**
	50	0.193	**0.189**	81.17	**81.56**	38.8	**38.2**
TR11	1	0.212	**0.201**	79.37	**80.39**	156.3	**140.7**
	10	0.191	**0.18**	81.45	**82.51**	154.3	**143.6**
	20	0.182	**0.177**	82.42	**82.87**	158.3	**122.2**
	50	0.174	**0.163**	83.19	**84.18**	152.5	**134.6**
Ionosphere	1	0.119	**0.097**	88.46	**90.57**	15.1	**7.33**
	10	0.102	**0.072**	90.03	**92.93**	12	**4.6**
	20	0.087	**0.072**	91.54	**92.99**	10.4	**6**
	50	0.077	**0.066**	92.48	**93.56**	8.2	**4.1**

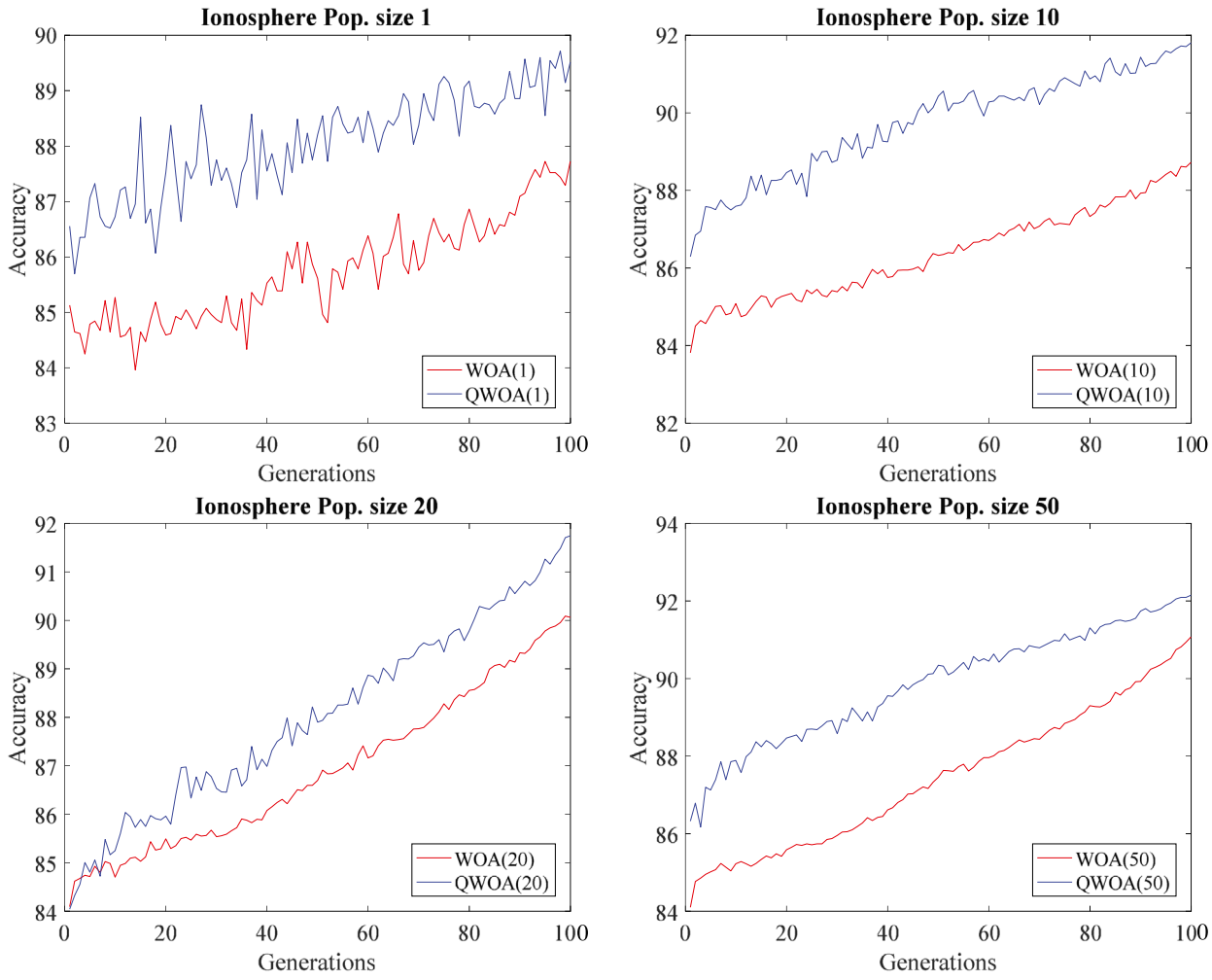

Fig. 3(a). Comparative plots for varying population sizes for Ionosphere.

The rotation angle (φ) and the direction of rotation are decided from a lookup table [36].

The proposed algorithm, QWOA for feature selection, and its flowchart are depicted in Figs. 1(a) and 1(b) respectively. Each whale position in a population of size n is represented in terms of d-dimensional Q-bits, corresponding to an individual. During *exploration*, the Tournament selection method is used to select a random individual that will guide the search. Crossover is performed between the mutated random individual and the mutated current individual. The *shrinking* method is carried out with the crossover between the best individual and its mutated form. The *spiral* movement is carried out using (1) and (2). As in WOA, the decision about the update method is based on the values of p and A. Eq. (6) is used to obtain the binary equivalent of the Q-bit individual. At each generation, for each quantum individual of the population, the quantum update is carried out for each Q-bit with the rotation gate using (10).

The fitness function, F used to evaluate the goodness of each binary feature vector computed as [28] :

$$F = \alpha\,(err) + \beta\,(k) \tag{11}$$

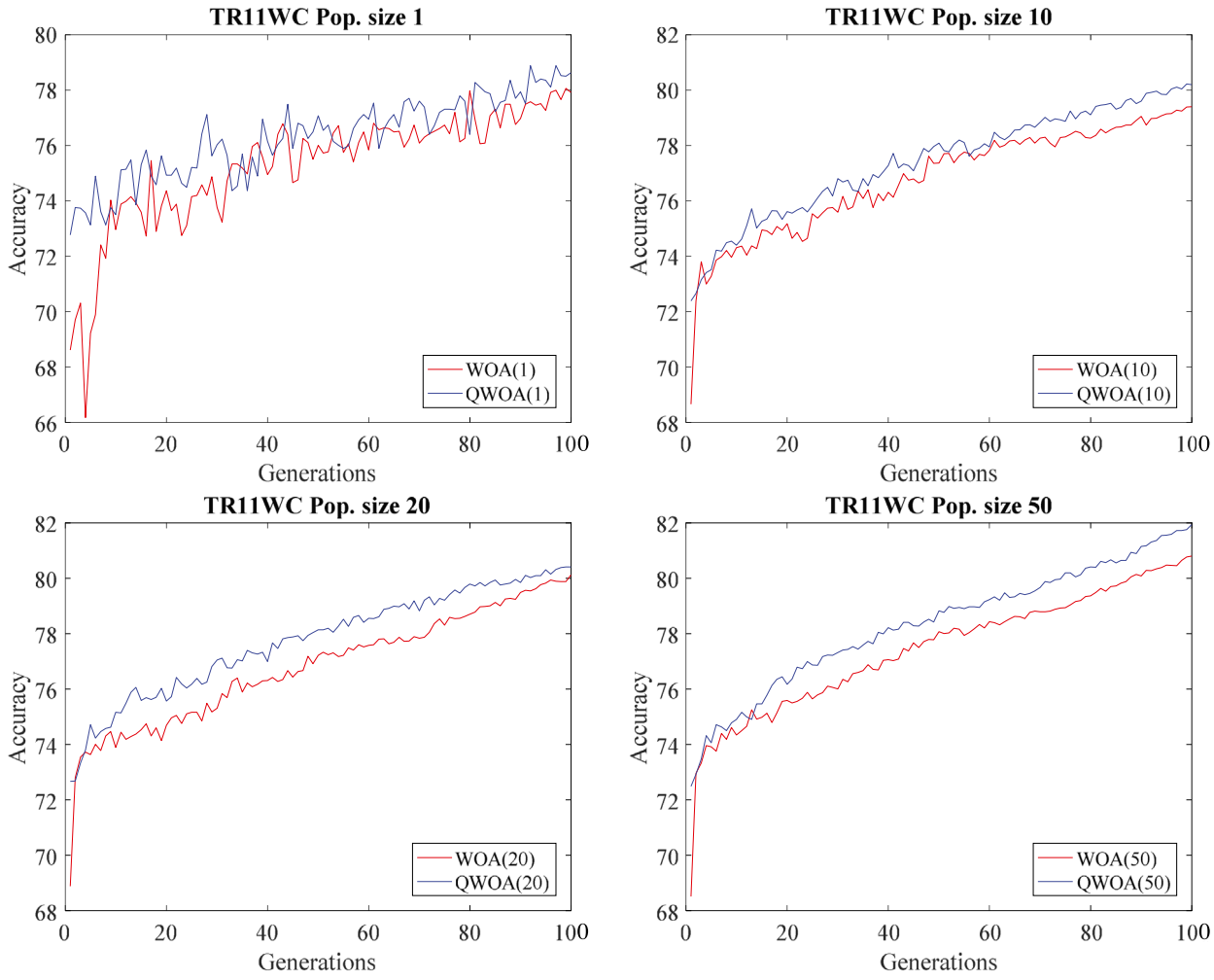

Fig. 3(b). Comparative plots for varying population sizes for TR11WC.

Where α and β give the weight to *err* and k respectively. It is designed to select that feature subset that minimizes the classification error (*err*) as well as the feature subset size (k). The fitness values at each generation are calculated and the whole population is updated using the shrinking, spiralling and exploration methods until the termination condition is reached.

3. Experimental section

For comparing the performance of the proposed QWOA method with the classical WOA and seven well-known meta-heuristic methods such as GA, PSO, BBA, GWO, SSA, QGA and QPSO, experiments have been performed on a set of fourteen challenging datasets from the domains of microarray gene expression, face image detection, high dimensional text and a collection of diverse datasets from the UCI repository. The datasets can be downloaded from [37–39]. The aim is to test the suitability of the proposed method by observing its impact across datasets and to be able to check if the performance improves across all the diverse datasets in comparison to the competing methods. The datasets are of low and high dimension depending upon the number of their features. Further, some datasets are two-class while others are multiclass in nature. As the datasets belong to diverse domains and the dimension of features of the datasets range from tens up-to thousands, the experiments performed form a sound basis to establish the suitability of the proposed method. The performance of the QWOA and other methods is evaluated in terms of the fitness function, classification accuracy, area under the ROC curve (AUC) and size of the selected features. We have performed the experiments with four well-known classifiers such as k-NN, LDC, SVM (linear kernel) and C4.5 The details of the datasets used are presented in Table 1.

The datasets are preprocessed, wherein m features with the highest variance are selected, and are normalized using z-score. For the global cancer map dataset (GCM), the preprocessing strategy used in the research work [40] is followed. For GCM, the train and test datasets were available separately.

In the case of high dimension datasets, if the whole feature set is given as input to a given meta-heuristic method, the required time and space will be very large. Moreover, as the number of samples available is very low, it leads to the problem of overfitting. Hence, it is desirable to build a system with a fewer number of features, which are relevant to classification. To achieve this, the experiments for high dimension datasets are performed in two phases. In the first phase, the input feature space is reduced through the Hierarchical Agglomerative Clustering method with Euclidean distance. This reduced feature set is then given as input to the method in the second phase.

The clustering in the first phase facilitates the grouping of features into a predetermined number of clusters. Further, from each of these clusters, a representative feature is selected using t-statistic, resulting in a smaller number of relevant and non-redundant features. The first phase of clustering greatly reduces the requirement of both time and space of the method in the

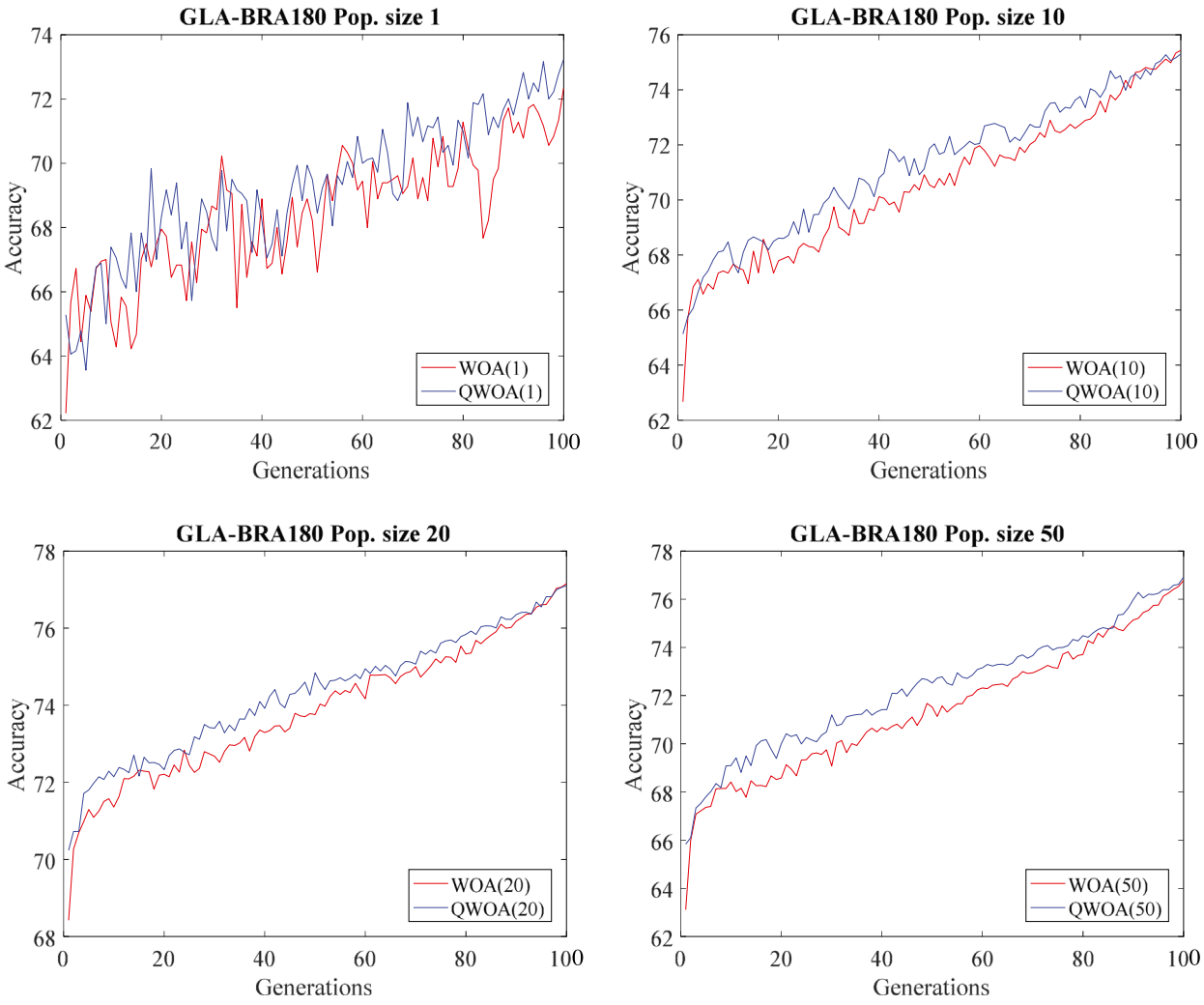

Fig. 3(c). Comparative plots for varying population sizes for GLA-BRA180.

Table 4
Wilcoxon signed rank test results.

QWOA vs. Methods	Fitness value					Classification accuracy					AUC				
	H at significance level α					H at significance level α					H at significance level α				
	p-value	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$	p-value	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$	p-value	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$
QWOA vs. GA	0.000122	1	1	1	1	0.000122	1	1	1	1	0.000244	1	1	1	1
QWOA vs. PSO	0.000244	1	1	1	1	0.000221	1	1	1	1	0.000221	1	1	1	1
QWOA vs. BBA	0.000221	1	1	1	1	0.000221	1	1	1	1	0.000221	1	1	1	1
QWOA vs. GWO	0.0004883	1	1	1	1	0.000122	1	1	1	1	0.000244	1	1	1	1
QWOA vs. SSA	0.0004883	1	1	1	1	0.000122	1	1	1	1	0.000244	1	1	1	1
QWOA vs. QGA	0.0002443	1	1	1	1	0.000122	1	1	1	1	0.00061	1	1	1	1
QWOA vs. QPSO	0.0001221	1	1	1	1	0.000122	1	1	1	1	0.000244	1	1	1	1
QWOA vs. WOA	0.0029	1	1	1	0	0.000122	1	1	1	1	0.0024	1	1	1	0

second phase. The reduced feature set, as shown in the last column of Table 1, is given as input to the second phase. The second phase involves the meta-heuristic method which further optimizes the feature set while improving the fitness value.

The population size for the experiments is fixed as 20, and the length of the individual is set as the feature set dimension as mentioned in Table 1. The experiments have been carried out for 100 generations. The 10-fold cross-validation scheme is used for evaluation, which is run 10 times. For each run, the fitness value of the best solution of the population is considered. The average fitness value from 10 runs is considered for comparison of the performance of the methods. Similarly, the average classification accuracy, the average number of selected features and the average AUC is calculated over 10 runs for comparison.

Tables 2(a)–2(d) show the average fitness value, the average classification accuracy, average AUC and the average number of selected features respectively with the k-NN classifier ($k = 5$). To investigate the stability and robustness of the proposed QWOA method in comparison to the other methods, the standard deviation is also reported for each dataset. The best performance

Table 5(a)
Comparison of classification accuracy and AUC with variation in classifier.

Dataset	Accuracy								AUC							
	KNN		LDA		SVM		C4.5		KNN		LDA		SVM		C4.5	
	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA
Lymphography	85.00	**86.01**	89.39	**89.73**	97.38	**97.62**	85.77	**86.10**	0.8282	**0.8520**	0.6525	**0.6812**	0.8452	**0.8519**	0.7897	**0.7931**
	1.18	1.11	0.64	0.53	4.39	3.42	0.53	0.79	0.0210	0.0209	0.0206	0.0470	0.0780	0.0801	0.0299	0.0259
Waveform	83.66	**83.87**	86.87	**86.91**	87.19	**89.03**	76.94	**77.65**	0.9496	**0.9511**	0.7223	**0.7263**	0.8672	**0.8903**	0.8935	**0.9021**
	0.42	0.27	0.14	0.08	0.52	1.61	0.34	0.35	0.0018	0.0024	0.0064	0.0034	0.5235	0.9464	0.0047	0.0049
Tictactoe	**89.56**	**89.56**	82.78	**82.80**	81.61	**81.79**	97.62	**97.72**	0.9732	**0.9737**	0.6184	**0.6236**	0.7816	**0.7845**	0.9863	**0.9865**
	0.00	0.00	0.09	0.08	2.84	2.74	0.20	0.17	0.0025	0.0022	0.0041	0.0044	0.0614	0.0632	0.0024	0.0024
StatlogHeart	86.37	**86.85**	86.85	**87.22**	92.96	**94.07**	85.22	**85.27**	0.8763	**0.8838**	0.7093	**0.7147**	0.8683	**0.8751**	0.8491	**0.8507**
	0.55	0.26	0.40	0.31	4.90	4.92	1.12	1.13	0.0156	0.0179	0.0083	0.0107	0.0687	0.0669	0.0116	0.0140
Ionosphere	91.82	**92.99**	89.63	**90.11**	97.14	**97.90**	94.81	**94.96**	0.9343	**0.9349**	0.5421	**0.5504**	0.8633	**0.8715**	**0.9330**	0.9329
	1.23	0.92	0.20	0.52	2.81	2.59	0.26	0.29	0.0141	0.0149	0.0085	0.0048	0.0851	0.0707	0.0103	0.0092
GLI-85	**100.00**	**100.00**	99.76	**100.00**	**100.00**	**100.00**	93.33	**93.80**	0.9996	**0.9998**	0.8174	**0.8463**	0.8944	**0.9356**	**0.8941**	0.8865
	0.00	0.00	0.50	0.00	0.00	0.00	0.78	0.75	0.0013	0.0012	0.0474	0.0484	0.1373	0.0800	0.0374	0.0352
GLA-BRA180	79.69	**80.88**	77.67	**78.17**	95.74	**97.04**	74.83	**74.93**	0.8998	**0.9010**	**0.7086**	0.6939	0.8761	**0.8827**	0.8169	**0.8242**
	1.15	1.11	1.79	1.34	4.30	4.06	1.09	1.18	0.0126	0.0006	0.0117	0.0310	0.0622	0.0439	0.0260	0.0252
GCM	71.52	**71.96**	73.26	**74.78**	68.91	**70.00**	64.35	**65.22**	0.8259	**0.8322**	0.9748	**0.9777**	0.8203	**0.8403**	**0.9708**	0.9671
	4.52	2.60	2.72	2.10	2.91	2.67	1.83	2.29	0.0300	0.0203	0.0134	0.0052	0.0225	0.0298	0.0141	0.0169
Tumour-9	75.30	**79.16**	82.50	**83.67**	98.33	**98.89**	59.11	**60.28**	0.7357	**0.7698**	0.8673	**0.8757**	0.8144	**0.8192**	0.8462	**0.8553**
	1.80	1.82	1.62	2.81	5.09	4.23	2.22	2.19	0.0151	0.0123	0.0273	0.0179	0.0946	0.1629	0.0317	0.0295
AR10P	99.46	**99.85**	99.77	**99.85**	**100.00**	**100.00**	76.15	**77.05**	0.9978	**0.9982**	0.9330	**0.9393**	**0.9401**	0.9399	0.8850	**0.8851**
	0.63	0.32	0.37	0.32	0.00	0.00	1.50	1.76	0.0025	0.0025	0.0175	0.0170	0.0360	0.0568	0.0188	0.0217
PIE10	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	87.73	**87.89**	**0.9998**	**0.9998**	0.9675	**0.9694**	0.9896	**0.9899**	**0.9318**	0.9316
	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.66	0.0002	0.0002	0.0079	0.0134	0.0143	0.0132	0.0114	0.0098
PIX10P	**99.90**	99.70	**100.00**	**100.00**	**100.00**	**100.00**	99.00	**99.03**	**0.9998**	0.9990	0.9920	**0.9982**	**0.9959**	0.9751	0.9852	**0.9854**
	0.32	0.48	0.00	0.00	0.00	0.00	0.00	0.18	0.0006	0.0026	0.0053	0.0035	0.0032	0.0143	0.0049	0.0047
TR11WC	82.42	**82.87**	68.57	**69.44**	95.37	**97.32**	79.94	**80.33**	0.9207	**0.9322**	0.6091	**0.6096**	0.9585	**0.9613**	0.9125	**0.9147**
	0.91	0.96	1.56	1.08	2.89	2.96	0.57	0.61	0.0083	0.0053	0.0092	0.0077	0.0132	0.0117	0.0080	0.0089
TR23WC	86.27	**86.37**	66.57	**67.40**	95.33	**97.33**	86.00	**86.45**	0.8866	**0.8917**	0.5408	**0.5540**	0.7239	**0.7246**	0.9381	**0.9422**
	1.77	1.80	1.44	1.21	4.90	3.14	0.79	0.87	0.0183	0.0204	0.0150	0.0129	0.0581	0.0719	0.0106	0.0081
Average	87.9275	**88.5766**	85.9734	**86.4344**	93.5699	**94.3565**	82.9143	**83.3350**	0.9162	**0.9228**	0.7611	**0.7686**	0.8742	**0.8816**	0.9023	**0.9041**

among all nine meta-heuristic methods for each of the fourteen datasets is shown in bold.

It can be observed from Table 2(a) that the performance of the QWOA method in terms of fitness value is best among all the 9 meta-heuristic methods, for each dataset. Hence the best average performance is achieved by the QWOA method. It can also be noted that WOA has the second best performance in terms of the overall average fitness value. From Table 2(b) we observe that QWOA outperforms other methods for 10 datasets, and achieves the same performance with some methods for 3 datasets. The QWOA achieved the second best classification accuracy for PIX10P dataset. However, the overall average of the QWOA method in terms of classification accuracy is best among the 9 competing methods. In Table 2(c), we observe that the QWOA method achieves the best AUC value for 12 datasets, same with another method for 1 dataset and second best with 1 dataset. The best overall average AUC is achieved by the QWOA method. Table 2(d) shows that no method is the clear winner among the 9 methods in terms of the average number of selected features. But it can be observed that the QWOA achieves the best performance followed by the QPSO and the QGA.

The deviation of each performance measure from the mean over different runs is obtained by the standard deviation. It can be observed from Tables 2(a)–2(d) that the standard deviation obtained for most of the methods is low and also comparable to each other. It can also be observed that the performance of the QWOA in terms of standard deviation of each performance measure is better than the WOA for most of the datasets. This suggests that the proposed QWOA provides consistent performance over different runs.

It can be observed from Tables 2(a)–2(d) that the performance of quantum algorithms QGA, QPSO and QWOA is better than their classical counterparts GA, PSO and WOA respectively in terms of average fitness value, average classification accuracy, average AUC and average number of selected features over all the considered datasets. The better performance of the quantum methods in comparison to their classical versions may be attributed to better representation of the population in terms of Q-bits and improved exploration and exploitation due to the involved operators.

Boxplot is a method to graphically represent the variation in the values through their central tendency measures. The higher the median/mean and the smaller the spread, the better is the method among methods in comparison. To understand the robustness of each method, for illustration, a boxplot of the classification accuracy using k-NN for the QWOA versus the boxplots of other methods are shown in Fig. 2 for (a) StatlogHeart (b) Ionosphere (c) GCM (d) Tumor_9 (e) AR10P and (f) TR11WC. It can be observed from the boxplots that both mean and median for the QWOA are higher than other competing methods for the StatlogHeart, Ionosphere, Tumor_9, AR10P and TR11WC datasets. The median of the WOA is higher than that of QWOA for the GCM dataset but the mean is comparable. Similar performance is also observed for other datasets.

To study the role of the population size on the performance of the classical and the quantum version of WOA, we conducted experiments with four different population sizes such as 1, 10, 20 and 50. Ten (10) independent runs of each algorithm were carried out for the 10 fold cross-validation with the k-Nearest Neighbour classifier. The maximum number of generations for each run is fixed to 100. Average results obtained with 10 runs for the three chosen datasets are presented in Table 3. GLA_BRA80 is a microarray dataset, TR11WC is a text-based dataset, while Ionosphere is a UCI dataset. These datasets are chosen from each of the domains to be able to explore diverse domains to show

Table 5(b)
Comparison of Fitness and Number of Selected Features with variation in classifier.

Dataset	Fitness								Number of Features							
	KNN		LDA		SVM		C4.5		KNN		LDA		SVM		C4.5	
	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA	WOA	QWOA
Lymphography	0.1538	**0.1445**	0.1137	**0.1115**	0.0283	**0.0262**	0.1462	**0.1431**	**9.6**	10.9	**15.7**	17.6	**4.2**	4.8	**9.5**	10.0
	0.0110	0.0109	0.0053	0.0052	0.0435	0.0342	0.0062	0.0077	2.7	1.4	1.9	0.5	1.9	1.9	3.6	3.7
Waveform	0.1692	**0.1662**	0.1379	**0.1370**	0.1333	**0.1153**	0.2353	**0.2283**	29.9	**25.9**	31.7	**29.6**	**26.2**	26.5	28.0	**26.6**
	0.0048	0.0032	0.0018	0.0010	0.0051	0.0160	0.0036	0.0034	3.5	2.7	3.1	1.8	1.8	3.5	4.9	5.0
Tictactoe	0.1089	**0.1085**	0.1754	**0.1752**	0.1856	**0.1839**	0.0311	**0.0297**	15.0	**14.0**	**13.3**	**13.3**	**9.7**	9.8	20.3	**19.4**
	0.0004	0.0000	0.0010	0.0009	0.0279	0.0271	0.0023	0.0017	1.2	0.0	0.5	0.5	1.5	1.3	3.1	2.2
StatlogHeart	0.1422	**0.1366**	0.1369	**0.1334**	0.0734	**0.0626**	0.1510	**0.1505**	9.4	**8.4**	**8.8**	9.0	**4.8**	5.1	6.2	**6.1**
	0.0065	0.0039	0.0044	0.0036	0.0484	0.0489	0.0129	0.0129	1.8	2.3	1.2	0.9	1.2	1.6	2.7	2.4
Ionosphere	0.0838	**0.0754**	0.1097	**0.1040**	0.0307	**0.0230**	0.0579	**0.0561**	9.6	**6.1**	24.0	**21.0**	8.4	**7.5**	22.2	**21.2**
	0.0135	0.0092	0.0027	0.0060	0.0277	0.0257	0.0028	0.0035	4.7	1.2	3.3	3.7	3.1	3.6	3.3	3.5
GLI-85	0.0033	**0.0028**	0.0049	**0.0031**	0.0007	**0.0003**	0.0715	**0.0669**	19.8	**17.0**	**15.7**	18.5	4.3	**2.0**	**32.9**	33.5
	0.0006	0.0007	0.0052	0.0006	0.0002	0.0002	0.0081	0.0081	3.5	4.1	5.3	3.6	1.3	0.9	7.9	10.7
GLA-BRA180	0.2080	**0.1960**	0.2267	**0.2215**	0.0455	**0.0331**	0.2555	**0.2549**	41.7	**40.3**	33.8	**32.0**	22.4	**19.1**	**37.8**	40.1
	0.0119	0.0118	0.0187	0.0137	0.0423	0.0403	0.0108	0.0115	8.0	6.5	9.0	3.9	8.8	5.4	8.6	4.3
GCM	0.2869	**0.2830**	0.2697	**0.2554**	0.3136	**0.3018**	0.3558	**0.3470**	32.3	**30.0**	**29.8**	34.4	35.2	**28.7**	16.8	**16.1**
	0.0447	0.0256	0.0274	0.0216	0.0293	0.0265	0.0180	0.0227	3.0	4.1	4.3	5.7	7.0	2.8	2.4	2.1
Tumour-9	0.2460	**0.2124**	0.1785	**0.1673**	0.0184	**0.0125**	0.4112	**0.4001**	43.3	**37.1**	**31.6**	33.4	11.2	**9.1**	**38.5**	40.8
	0.0188	0.0183	0.0157	0.0271	0.0503	0.0417	0.0221	0.0219	7.2	2.9	7.8	6.5	2.5	3.9	7.3	6.1
AR10P	0.0112	**0.0074**	0.0078	**0.0070**	0.0019	**0.0015**	0.2433	**0.2342**	59.0	**58.3**	55.4	**54.4**	19.0	**14.9**	72.1	**70.1**
	0.0064	0.0040	0.0041	0.0031	0.0003	0.0005	0.0153	0.0174	5.9	10.0	7.1	5.8	2.9	5.1	13.0	9.1
PIE10	0.0041	**0.0034**	0.0026	**0.0025**	**0.0006**	0.0007	0.1285	**0.1269**	40.6	**36.2**	26.3	**25.4**	12.6	**7.1**	70.3	**69.9**
	0.0007	0.0007	0.0003	0.0002	0.0002	0.0002	0.0070	0.0069	6.9	7.3	2.8	2.5	2.1	2.0	11.0	9.4
PIX10P	0.0048	**0.0043**	0.0017	**0.0015**	0.0011	**0.0003**	0.0136	**0.0114**	18.3	12.8	17.4	**14.5**	10.6	**3.0**	36.7	**18.6**
	0.0030	0.0049	0.0003	0.0005	0.0003	0.0001	0.0005	0.0021	3.1	3.4	2.9	5.2	2.7	1.0	4.8	13.4
TR11WC	0.1820	**0.1769**	0.3168	**0.3083**	0.0497	**0.0309**	0.2061	**0.2017**	158.3	**122.2**	**114.4**	115.7	75.4	**72.4**	150.2	**139.8**
	0.0096	0.0099	0.0155	0.0108	0.0287	0.0294	0.0062	0.0064	19.2	29.1	11.4	13.4	14.2	11.5	28.0	15.6
TR23WC	0.1422	**0.1404**	0.3355	**0.3263**	0.0482	**0.0296**	0.1462	**0.1411**	126.8	**110.7**	89.8	**72.4**	**40.7**	64.7	152.3	**139.9**
	0.0180	0.0181	0.0143	0.0117	0.0485	0.0311	0.0079	0.0088	17.3	22.4	5.6	12.4	13.7	9.2	25.1	14.3
Average	0.1247	**0.1184**	0.1441	**0.1396**	0.0665	**0.0587**	0.1752	**0.1709**	43.8	**37.9**	36.3	**35.1**	20.3	**19.6**	49.6	**46.6**

Table 6
Summary of performance of QWOA with different classifiers on the basis of classification accuracy, AUC, fitness value and number of features based on Tables 5(a) and 5(b).

	Classification accuracy			AUC			Fitness value			Number of features		
	Wins	Draws	Trails	Wins	Draws	Trails	Wins	Draws	Trails	Wins	Draws	Trails
k-NN	10	3	1	12	1	1	14	0	0	12	1	1
LDA	12	2	0	13	1	0	14	0	0	13	0	1
SVM	10	4	0	12	2	0	13	0	1	12	0	2
C4.5	14	0	0	10	4	0	14	0	0	10	0	4

the comparison between the QWOA and the WOA. Based on the comparative results, the following can be observed:

- For each of the population sizes, the QWOA outperforms the WOA for all the datasets in terms of the average value of the three chosen metrics.
- Both the average fitness value and the average classification accuracy improve with the increase in the population size for all the datasets.

Figs. 3(a)–3(c) plot the comparison of the growth of the average accuracy between the QWOA and the WOA over 100 generations for the three datasets for the population size 1, 10, 20 and 50. It is clearly evident that the QWOA outperforms the WOA consistently for every population size, across generations. In general, initially, the average performance of the QWOA is better than that of WOA, as the generations advance, the results become similar. Even with the population of a single individual, the QWOA performs better than WOA. The QWOA explores and exploits the search space better due to its quantum representation, thereby giving better performance of the QWOA than that of the WOA. The consistent improvement of performance, for all the chosen population sizes, point towards the potential of the quantum version of the WOA as a possible metaheuristic algorithm of swarm computing.

To determine whether the QWOA and other eight meta-heuristic methods are statistically significantly different, the non-parametric two-sided Wilcoxon Signed Rank test has been carried out between the QWOA and each of the eight methods. The test does not assume a normal distribution, and outliers do not affect its performance [49]. It tests the presence of a significant statistical difference between the two methods. The statistical significant difference between each pair of methods is evaluated on the basis of the fitness value, classification accuracy and AUC on the 14 datasets. Their p-values are tabulated in Table 4. Based on the decision test, the logical value, h=1, indicates a statistically significant difference between the two methods at the given significance level. For the levels of significance, $\alpha = 0.05$, $\alpha = 0.01$ and $\alpha = 0.005$, the QWOA outperformed all the eight methods in terms of all the three metrics obtained using the *k*-NN classifier, viz. fitness value, classification accuracy, and AUC. For $\alpha = 0.001$, the QWOA is statistically significantly better than all other methods for all the three metrics except with the WOA for the fitness value and AUC.

Table 7
Comparison of the proposed QWOA with state of the art.

Dimension	Dataset	QWOA	Literature	QWOA	Literature	Ref.
		Accuracy		Feature %		
High	GLI-85	**100.000**	0.980	–	–	[41]
High	GLA_BRA180	**80.89**	77.890	**0.001**	0.030	[42]
High	GCM	**71.96**	52.170	**0.00002**	0.00003	[43]
High	Tumor-9[a]	**87.600**	85.19	**0.005**	16.000	[44]
High	AR10P	**99.840**	98.000	**0.024**	0.063	[45]
High	PIE10P	**100.000**	98.670	**0.015**	0.167	[46]
High	PIX10P	**99.700**	99.000	**0.001**	0.005	[42]
High	TR11WC[b]	**84.200**	84.110	**0.022**	0.160	[42]
High	TR23WC	**97.619**	97.250	–	–	[46]
Low	Lymphography	**86.010**	85.000	**0.606**	0.533	WOA
Low	Waveform	**83.866**	83.66	**0.648**	0.748	WOA
Low	Tictactoe	**89.562**	80.000	–	–	[47]
Low	StatlogHeart	**86.850**	84.380	–	–	[48]
Low	Ionosphere	**92.99**	83.340	–	–	[48]

[a]SVM classifier.
[b]C4.5 classifier.

To understand the variation in the performance of the proposed method with the choice of the classifier, we have investigated four classifiers, such as k-NN, LDA, SVM, and C4.5. The performance is compared with the WOA for each classifier. The comparison of the QWOA with the WOA in the terms classification accuracy and AUC with variation in the classifier is shown in Table 5(a). Similarly, the comparison of the QWOA with the WOA in terms of fitness value and the number of selected features is shown in Table 5(b). The best performance for each classifier for each dataset is shown in bold. A summary of observations regarding the performance of the different classifiers for the metrics: classification accuracy, AUC, fitness value and the minimum number of features is shown in Table 6. The summary is shown in terms of the number of wins, draws, and losses. It can be observed from Tables 5(a) and 5(b) that the average of the classification accuracy, AUC, fitness values, and the minimum number of selected features obtained using QWOA is superior to the WOA for all the four classifiers.

The proposed method QWOA is compared with the state-of-the-art in Table 7, both on the basis of classification accuracy (*k*-Nearest Neighbour classifier) and also on the basis of the percentage ratio of the selected feature size. In the case of Tumor-9, linear SVM is used for evaluating the performance as described in [44,50]. Similarly, for TR23WC, the C4.5 classifier is used as suggested in the work of Wang et al. [46]. It can be observed from Table 7 that the QWOA has consistently outperformed all the state-of-the-art methods for all the datasets. Experiments on datasets from diverse domains for both low and high dimensional features have shown the superior performance of the QWOA over the existing methods. The QWOA has also performed better for both the binary-class and the multi-class datasets in comparison to the existing methods. Moreover, the selected feature % is also very low, implying the low cost of building the decision model. The QWOA extends the promise of high performing yet low complexity wrapper based nature inspired feature selection method.

4. Conclusion

In this paper, a quantum-based Whale Optimization Algorithm (QWOA) is proposed that enhances the diversification and convergence properties of the classical Whale Optimization Algorithm for feature selection. The quantum representation of whale positions, in the QWOA facilitates a better chance of population diversity, and hence provide better exploration even with smaller population size. Modified mutation and crossover operators are introduced for quantum-based exploration, shrinking and spiral movement of the whales in the proposed QWOA. The application of the quantum rotation gate operator balances the exploration (diversification) and exploitation (convergence) property of the proposed method. Further, reduction of the feature input set using the clustering step prior to QWOA for high dimensional datasets is also instrumental in building a high performing decision system. Experiments on fourteen diversified datasets establish the merit of the proposed method as a meta-heuristic method that performs well on challenging datasets of both low and high feature dimensions. Comparison of the QWOA with eight meta-heuristic methods shows the improved performance in terms of the average fitness, average classification accuracy and average AUC. The statistical test also demonstrates the significantly better performance of the QWOA in comparison to the other competing methods. Experimental results demonstrate that the performance of quantum algorithms QGA, QPSO and QWOA is better than their classical counterparts GA, PSO and WOA respectively in terms of average fitness value, average classification accuracy, average AUC and average number of selected features over all the considered datasets. The better performance of the quantum methods in comparison to their classical versions may be attributed to better representation of the population in terms of Q-bits and improved exploration and exploitation due to the involved operators.

In future work, population selection methods other than the Tournament method can be explored to improve the performance further. Quantum variation operators like the Hadamard, the NOT gate and the controlled-NOT gate, and other lookup tables can also be tested for faster convergence. The proposed method is effective only for feature selection or binary unconstrained optimization problem. Further modification in the QWOA is required to solve continuous optimization problems and multi-objective optimization problems, which we will explore in future.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106092.

CRediT authorship contribution statement

R.K. Agrawal: Conceptualization, Methodology. **Baljeet Kaur:** Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Surbhi Sharma:** Validation, Visualization.

Acknowledgment

Author Baljeet Kaur would like to acknowlege the support and funding by Hansraj College, University of Delhi to pursue her post-doctoral research.

References

[1] J.H. Holland, Genetic algorithms, Sci. Am. 267 (1) (1992) 66–73.
[2] J.R. Koza, Genetic Programming, 1997.
[3] K.-S. Tang, K.-F. Man, S. Kwong, Q. He, Genetic algorithms and their applications, IEEE Signal Process. Mag. 13 (6) (1996) 22–37.
[4] U. Mehboob, J. Qadir, S. Ali, A. Vasilakos, Genetic algorithms in wireless networking: Techniques, applications, and issues, Soft Comput. 20 (6) (2016) 2467–2501.
[5] R. Aguilar-Rivera, M. Valenzuela-Rendón, J. Rodríguez-Ortiz, Genetic algorithms and Darwinian approaches in financial applications: A survey, Expert Syst. Appl. 42 (21) (2015) 7684–7697.
[6] M. Dorigo, G. Di Caro, Ant colony optimization: A new meta-heuristic, in: Presented at the Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), vol. 2 1999, pp. 1470–1477.
[7] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the 1995 IEEE International Conference on Neural Networks, Vol. 4, Perth, Australia, IEEE Service Center, Piscataway, NJ, 1995, p. 1942.
[8] B. Basturk, An artificial bee colony (ABC) algorithm for numeric function optimization, in: presented at the IEEE Swarm Intelligence Symposium, Indianapolis, IN, USA, 2006, pp. 2006.
[9] R.Y.M. Nakamura, L.A.M. Pereira, D. Rodrigues, K.A.P. Costa, J.P. Papa, X.-S. Yang, Binary bat algorithm for feature selection, in: Swarm Intelligence and Bio-Inspired Computation, Elsevier, 2013, pp. 225–237.
[10] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[11] H. Faris, et al., An efficient binary salp swarm algorithm with crossover scheme for feature selection problems, Knowl.-Based Syst. 154 (2018) 43–67.
[12] W. Deng, H. Zhao, X. Yang, J. Xiong, M. Sun, B. Li, Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment, Appl. Soft Comput. 59 (2017) 288–302.
[13] Y. Chen, Y. Lin, Controlling the movement of crowds in computer graphics by using the mechanism of particle swarm optimization, Appl. Soft Comput. 9 (3) (2009) 1170–1176.
[14] M.M. al-Rifaie, A. Aber, R. Sayers, E. Choke, M. Bown, Deploying swarm intelligence in medical imaging, in: Presented at the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2014, pp. 14–21.
[15] X. Yao, Y. Liu, G. Lin, Evolutionary programming made faster, IEEE Trans. Evol. Comput. 3 (2) (1999) 82–102.
[16] R. Storn, K. Price, Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359.
[17] N. Hansen, S.D. Müller, P. Koumoutsakos, Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), Evol. Comput. 11 (1) (2003) 1–18.
[18] X.-S. Yang, Bat algorithm for multi-objective optimisation, 2012, arXiv preprint arXiv:1203.6571.
[19] S. Mirjalili, S. Saremi, S.M. Mirjalili, L. dos S. Coelho, Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization, Expert Syst. Appl. 47 (2016) 106–119.
[20] I.-S. Oh, J.-S. Lee, B.-R. Moon, Hybrid genetic algorithms for feature selection, IEEE Trans. Pattern Anal. Mach. Intell. (11) (2004) 1424–1437.
[21] J. Yang, V. Honavar, Feature subset selection using a genetic algorithm, in: Feature Extraction, Construction and Selection, Springer, 1998, pp. 117–136.
[22] C.-L. Huang, J.-F. Dun, A distributed PSO–SVM hybrid system with feature selection and parameter optimization, Appl. Soft Comput. 8 (4) (2008) 1381–1391.
[23] B. Xue, M. Zhang, W.N. Browne, Particle swarm optimization for feature selection in classification: A multi-objective approach, IEEE Trans. Cybern. 43 (6) (2013) 1656–1671.
[24] S. Kashef, H. Nezamabadi-pour, An advanced ACO algorithm for feature subset selection, Neurocomputing 147 (2015) 271–279.
[25] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, Neurocomputing 172 (2016) 371–381.
[26] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (1–3) (2002) 389–422.
[27] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1–2) (1997) 273–324.
[28] M. Mafarja, S. Mirjalili, Whale optimization approaches for wrapper feature selection, Appl. Soft Comput. 62 (2018) 441–453.
[29] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.
[30] K.-H. Han, J.-H. Kim, Quantum-inspired evolutionary algorithm for a class of combinatorial optimization, IEEE Trans. Evol. Comput. 6 (6) (2002) 580–593.
[31] A. Narayanan, M. Moore, Quantum-inspired genetic algorithms, in: Presented at the Evolutionary Computation, 1996. Proceedings of IEEE International Conference on, 1996, pp. 61–66.
[32] M. Sardana, R. Agrawal, B. Kaur, Clustering in conjunction with quantum genetic algorithm for relevant genes selection for cancer microarray data, in: presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2013, pp. 428–439.
[33] M. Xi, J. Sun, L. Liu, F. Fan, X. Wu, Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine, Comput. Math. Methods Med. (2016) 2016.
[34] J. Sun, B. Feng, W. Xu, Particle swarm optimization with particles having quantum behavior, in: Presented at the Proceedings of the 2004 congress on evolutionary computation (IEEE Cat. No. 04TH8753), vol. 1, 2004, pp. 325–331.
[35] M. Mitchell, An Introduction to Genetic Algorithms, MIT press, 1998.
[36] A. Abderrahim, E.-G. Talbi, M. Khaled, Hybridization of genetic and quantum algorithm for gene selection and classification of microarray data, Internat. J. Found Comput. Sci. 23 (02) (2012) 431–444.
[37] ASU datasets. [Online]. Available: http://featureselection.asu.edu/old/datasets.php.
[38] Text Datasets. [Online]. Available: http://tunedit.org/repo/Data/Text-wc.
[39] UCI datasets. [Online]. Available: https://archive.ics.uci.edu/ml/index.php.
[40] S. Ramaswamy, et al., Multiclass cancer diagnosis using tumor gene expression signatures, Proc. Natl. Acad. Sci. 98 (26) (2001) 15149–15154.
[41] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, J.M. Benítez, F. Herrera, A review of microarray datasets and applied feature selection methods, Inform. Sci. 282 (2014) 111–135.
[42] Q. Song, J. Ni, G. Wang, A fast clustering-based feature subset selection algorithm for high-dimensional data, IEEE Trans. Knowl. Data Eng. 25 (1) (2011) 1–14.
[43] M. Sardana, R. Agrawal, B. Kaur, An incremental feature selection approach based on scatter matrices for classification of cancer microarray data, Int. J. Comput. Math. 92 (2) (2015) 277–295.
[44] A.J. Ferreira, M.A. Figueiredo, An unsupervised approach to feature discretization and selection, Pattern Recognit. 45 (9) (2012) 3048–3060.
[45] N. Bi, J. Tan, J.-H. Lai, C.Y. Suen, High-dimensional supervised feature selection via optimized kernel mutual information, Expert Syst. Appl. 108 (2018) 81–95.
[46] G. Wang, Q. Song, B. Xu, Y. Zhou, Selecting feature subset for high dimensional data via the propositional FOIL rules, Pattern Recognit. 46 (1) (2013) 199–214.
[47] H. Faris, S. Mirjalili, I. Aljarah, M. Mafarja, A.A. Heidari, Salp Swarm algorithm: Theory, literature review, and application in extreme learning machines, in: Nature-Inspired Optimizers, Springer, 2020, pp. 185–199.
[48] Y. Gu, L. Cheng, Classification of class overlapping datasets by kernel-MTS method, Int. J. Innov. Comput. Inform. Control 13 (5) (2017) 1759–1767.
[49] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18.
[50] J.E. Staunton, et al., Chemosensitivity prediction by transcriptional profiling, Proc. Natl. Acad. Sci. 98 (19) (2001) 10787–10792.