An Enhanced Quantum-Inspired Evolutionary Fuzzy Clustering

Neha Bharill  
Department of Computer Science and Engineering  
Indian Institute of Technology  
Indore, India 453331  
Email: phd12120103@iiti.ac.in

Om Prakash Patel  
Department of Computer Science and Engineering  
Indian Institute of Technology  
Indore, India 453331  
Email: phd1301201003@iiti.ac.in

Aruna Tiwari  
Department of Computer Science and Engineering  
Indian Institute of Technology  
Indore, India 453331  
Email: artiwari@iiti.ac.in

Abstract—Clustering is one of the widely used knowledge discovery techniques to reveal the structures in a dataset that can be extremely useful for the analyst. In fuzzy based clustering algorithms, the procedure acquired for choosing the fuzziness parameter \( m \), the number of clusters \( C \) and the initial cluster centroids \( V_C \) is extremely important as it has a direct impact on the formation of final clusters. Moreover, the improper selection of these parameters may lead the algorithms to the local optima. In this paper, we proposed an Enhanced Quantum-Inspired Evolutionary Fuzzy C-Means (EQIE-FCM) algorithm to compute the global optimal value of these parameters. In EQIE-FCM, we utilize the quantum computing concept in combination with fuzzy clustering to evolve the different values of these parameters in several generations. However, in each generation these parameters are represented in terms of a quantum bit \( \psi \). At each generation \( \psi \), the quantum bit of these parameters is updated using a quantum rotational gate. Through this, after several generations of evolution, we get the global optimal values of these parameters from a large quantum search space. The EQIE-FCM algorithm is applied on the Pima Indians Diabetes dataset and the performance of EQIE-FCM is compared with another Quantum-inspired Fuzzy Clustering (QIE-FCM) and other three fuzzy based evolutionary clustering algorithms from the literature. Extensive experiments indicate that the EQIE-FCM algorithm outperforms many baseline approaches and can be used an effective clustering algorithm.

I. INTRODUCTION

Nowadays, data mining techniques are gaining importance in the medical research in order to analyze the large volume of medical data. Clustering is one of the most widely used data mining technique. The main goal of clustering is to divide the data set into groups such that the intra-cluster similarity is maximized, and inter-cluster similarity is minimized. This signifies that a cluster is a collection of data points such that the data points lying within the same cluster are more similar to each other than to the data points lying in the other clusters. Clustering is also referred as an unsupervised learning approach because it is used as an important tool for finding the hidden patterns and structures from a large database without the background knowledge. Clustering algorithms are broadly classified as hierarchical and partitional clustering [1], [2]. Hierarchical clustering groups the data points with the sequence of partitions, either from singleton clusters to a cluster including all individuals or vice versa. Partitional clustering algorithm attempts to divide \( N \) data points into \( C \) number of clusters and produce \( C \) fuzzy partitions that optimize a criteria function. Recently, partitional clustering algorithms have been widely adopted by the researchers due to the linear time complexity and low computational requirements [3].

Fuzzy C-Means algorithm is one of the most widely used partitional clustering algorithms was initially given by Dunn [4] and generalized by Bezdek [5]. FCM partitions a collection of \( N \) data points \( X = \{x_1, ..., x_N\} \) into \( C \) fuzzy clusters such that a cluster centroid corresponding to each cluster is obtained by minimizing a criterion function of dissimilarity measure. FCM algorithm employs fuzzy partitioning such that a data point can belong to the several clusters with a membership degree \( U = \{\mu_{i}\} \) which is allowed to take any value between 0 and 1. This membership value indicates the degree to which the point is more representative of one cluster than the other. Even though with the fuzzy clustering the accuracy of cluster representation increases, but there are several fundamental sources of ambiguity in clustering. One of the major issues with the FCM algorithm is that the number of clusters in a dataset has to be specified in advance. This is because to perform clustering of different datasets, different number of clusters are required, which is difficult to be known beforehand. The second problem is to decide what initial cluster centroids are to be used to form clusters. Generally, the FCM algorithm starts with the random assignment of cluster centroids as the initialization process. The behaviour of FCM algorithm is highly dependent upon the selection of initial cluster centers and always converges to the nearest local optima from the starting position of the search. However, FCM does not guarantee unique clustering because we get different results with randomly chosen initial centers. Due to this, the clustering results generated by the FCM algorithm produces inconsistent results. Thus, the final cluster centers may not be the optimal ones as the algorithm converges to the local optimal solutions. Another source of ambiguity in FCM algorithm is the selection of an appropriate value of fuzziness parameter \( m \) for a dataset because it widely varies from one dataset to another. The choice of inappropriate value of \( m \) may also lead the FCM algorithm to the local optima problem.

In order to overcome the disadvantages manifested above, the researchers from diverse fields are applying cluster validity index [6], [7] and evolutionary fuzzy based algorithms inspired...
by the concept of quantum computing in many application areas like distributed computing [8], image segmentation [9] and control system [10]. In addition to this, some evolutionary clustering algorithms are proposed in combination with genetic algorithm [11] and differential evolutionary [12] to overcome the problem of local optima. Furthermore, Karelowski and Vidya [13] proposed an approach for clustering diabetes data by applying genetic algorithm in combination with entropy based fuzzy clustering to find the initial cluster centroids. Chaoshun and Jianzhong [14], proposed a new fuzzy clustering algorithm based on chaos optimization which combines mutative scale chaos optimization, strategy and gradient method together. It optimizes the clustering objective function and performs clustering automatically without knowing the number of clusters in advance. Palanisamy and Selvan [15] proposed a novel method named as entropy-based fuzzy clustering to identify the relevant subspaces in the functional workspace. In this approach, a heuristic method based on the Silhouette criterion was used to find the number of clusters. In spite of the wide popularity of above stated approaches and applications, it is still a challenging issue to decide all the initialization parameters of FCM algorithm. Therefore, finding the appropriate value of $m$, $C$ and the initial cluster centroids are the key aspects for eliminating the local convergence problem of FCM algorithm.

Considering the shortcomings of FCM algorithm, in this paper an Enhanced Quantum-Inspired Evolutionary Fuzzy C-Means (EQIE-FCM) algorithm is proposed. In this approach, we utilize the concept of quantum computing in combination with fuzzy clustering for evolving the fuzziness parameter $m$, the number of clusters $C$ and the initial cluster centers in several generations. In EQIE-FCM algorithm, we adopted $\text{VI}_{D_{SO}}$ index as the objective function to evaluate the fitness of produced partitions in each generation ($g$). After several generations of evolution, we guarantee to achieve the global optimal value of these parameters from a large quantum search space. We perform a group of experiments to validate the performance of EQIE-FCM algorithm in comparison with QIE-FCM algorithm [16]. The QIE-FCM algorithm is also a quantum based fuzzy clustering approach aims to eliminate the problem of local convergence in FCM. However, it is able to find the global optimal value of $m$ from a large quantum search space by representing the parameter in terms of qubits. But, due to the random selection of value of $C$ in this approach, it may again trap into the problem of local optima. Thus, in EQIE-FCM, we aim to find the global optimal value of both the parameters by representing these parameters in terms of qubits which provide better characteristic of population diversity than other representation [20]. In addition to this, to validate the efficacy of EQIE-FCM algorithm, we compared it with the other evolutionary fuzzy based clustering algorithms [8], [9], [11]. We evaluate the performance of these approaches on well known Pima Indians Diabetes data available on UCI Machine Learning Repository [17]. The experimental results prove the efficacy of EQIE-FCM algorithm in terms of finding the global optimal value of $m$, $C$ and initial cluster centroids.

The rest of the paper has been structured as follows: Section II brief description of Pima Indian Diabetes dataset. Section III briefly explains the concept of quantum computing. The detailed discussion of the proposed algorithm is presented in Section IV. The experimental setup and analysis with experimental results on Pima Indians Diabetes data is presented in Section V. Finally, Section VII present the concluding remarks.

## II. DESCRIPTION OF PIMA INDIANS DIABETES DATA

Diabetes is one of the world’s most prevalent chronic disease that occurs when the pancreas is unable to produce enough insulin or when the body cells cannot utilize the produced insulin. It is becoming a common crisis among the majority of adults in developed countries and are increasing rapidly in developing countries. This is because of the rapid advancement in the technology which has brought a significant change in the lifestyle and eating habits. People are getting more prone to the wide range of fast foods and ready-to-eat processed food promoting by multinational companies. Due to unhealthy eating habits and intake of excessive calories, it is a major driving force behind escalating obesity and overweight worldwide. The overweight and obesity are driving the global diabetes epidemic. Diabetes is broadly categories into two types referred as type 1 diabetes and type 2 diabetes. Type 1 diabetes is caused due to the lack of insulin production in the body, and it is commonly seen in the children and young adults under the age of 40 years. Conversely, type 2 diabetes is the most common form of diabetes, which occurs because the body cells is unable to utilize the produced insulin. Worldwide, approximately 10% of the patient is suffering from type 1 diabetes and rest 90% are suffering from type 2 diabetes. According to the World Health Organization in 2014, it is estimated that over 347 million people throughout the world had diabetes, and the figure is expected to rise to 330 million by 2025 out of which 52 million people will be Indians, largely due to population growth, unhealthy eating habits and a sedentary lifestyle [18], [19]. In this study, we performed our experiment on Pima Indians Diabetes (PID) dataset availed from UCI Machine Learning Repository [17]. The dataset comprises of two categories, i.e. “Tested positive” which involves 65.10% of the dataset (500 samples) whilst “Tested negative” involves 34.89% of the dataset (268 samples) where each sample consists of 8 numerical features. The detailed description of each feature is given in Table I and Fig. 1 represent the scatter plot of Pima Indian Diabetes data in two-dimensional space.

<table>
<thead>
<tr>
<th>ID</th>
<th>Attribute</th>
<th>ID</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No. of times pregnant (NTP)</td>
<td>5</td>
<td>Diastolic Blood pressure in mmHg (DBP)</td>
</tr>
<tr>
<td>2</td>
<td>Plasma glucose concentration (PGC)</td>
<td>6</td>
<td>Body mass index in kg/m^2 (BMI)</td>
</tr>
<tr>
<td>3</td>
<td>2-h serum insulin in mU/ml (SI)</td>
<td>7</td>
<td>Diabetes pedigree function (DPF)</td>
</tr>
<tr>
<td>4</td>
<td>Triceps skin fold thickness in mm (TSF)</td>
<td>8</td>
<td>Years of age (YDA)</td>
</tr>
</tbody>
</table>

![Fig. 1. Scatter plot of Pima Indians diabetes data in two dimensional space reflecting the presence of three clusters marked with a circle.](image)
III. PRELIMINARIES

Quantum computing concept represents the data in terms of quantum bits ($Q$). In general, a quantum bit consists of several qubits ($q_p$) and can be represented as follows:

$$Q = (q_1, q_2, \ldots, q_K) \quad (1)$$

Where, $p = 1, 2, 3, \ldots, K$ and $K$ shows the number of qubits to form a quantum bit ($Q$). Qubits ($q_p$) are a smallest unit of information representation. Generally, qubits differ from the classical computer bits in terms of representation and storage. As a classical bit represents only two possibilities of any event at one time by bit “1” or “0”. However, a qubit can exist in both states simultaneously using the probability concept proposed by Han and Kim [20], [21]. Qubit shows the linear superposition of “1” and “0” bits probabilistically, which is denoted as follows:

$$q_p = \alpha_p |0\rangle + \beta_p |1\rangle \quad (2)$$

where and $\alpha$, $\beta$ are the complex numbers representing the probability of qubit in “1” state and in “0” state. A probability model is applied here, which represent “0” state by $\alpha_p^2$ and “1” state by $\beta_p^2$, where

$$\alpha_p^2 + \beta_p^2 = 1; 0 \leq \alpha_p \leq 1, 0 \leq \beta_p \leq 1 \quad (3)$$

As shown above, a quantum bit ($Q$) formed by a single qubit $q_p$ where $p = 1$ can represent two states, e.g. “0” or “1” state. Similarly, a quantum bit ($Q$) consists of two-qubits i.e. $q_p$ where $p = 1, 2$ can represent the linear superposition with four states i.e. “00”, “01”, “10” and “11”. Here is an example that explains the essence of a quantum bit formed using two-qubits are represented as follows:

$$Q = \left\langle \alpha_1 |\alpha_2 \right\rangle \beta_1 |\beta_2 \rangle \quad (4)$$

As mentioned in Eq (3) that the value of $\alpha$ and $\beta$ lies in the range of “0” and “1”. Therefore, $\alpha$ and $\beta$ can be initialized with any value in the above mentioned interval as follows:

$$Q = \left\langle \frac{1}{\sqrt{2}} |\frac{1}{\sqrt{2}} \right\rangle \quad (5)$$

A quantum bit ($Q$) formulated in terms of two-qubits consists of 4 different stages, which are represented as follows:

$$Q = (\alpha_1 \times \alpha_2)(00) + (\alpha_1 \times \beta_2)(01)$$
$$+ (\alpha_2 \times \beta_1)(10) + (\beta_1 \times \beta_2)(11) \quad (6)$$

$$Q = (1/ \sqrt{2}) \times (1/ \sqrt{2})(00) + (1/ \sqrt{2}) \times (1/ \sqrt{2})(01)$$
$$+ (1/ \sqrt{2}) \times (1/ \sqrt{2})(10) + (1/ \sqrt{2}) \times (1/ \sqrt{2})(11) \quad (7)$$

Similar to the above mentioned equation, a quantum bit ($Q$) formed by $K$ qubits can represent $2^K$ states at the same time. As we can see in Eq (7), single quantum bit ($Q$) is enough to represent four states. Thus, the quantum bit ($Q$) representation provides better characteristics of population diversity in comparison with other representations and also enables us to find the global optimal solution from the large search space. Han and Kim [20] use the concept of quantum computing in combination with genetic algorithm for evolving the optimal solution of the knapsack problem in several generations. Based on the above aforementioned idea, we proposed EQIE-FCM algorithm, which uses the quantum computing concept in combination with fuzzy clustering. The proposed approach finds the global best value of the fuzziness parameter $m$ and the number of clusters $C$ with the best location of initial cluster centroids for Pima Indians Diabetes data from a large quantum search space in several generations.

IV. PROPOSED APPROACH

As suggested by Pal and Bezdek [6], the fuzziness parameter $m$ and the number of cluster $C$ play a major to validate the fitness of partitions produced by fuzzy based clustering algorithms. In this study, we proposed an Enhanced Quantum-Inspired Evolutionary Fuzzy C-Means (EQIE-FCM) Algorithm to investigate the appropriate value of these parameters for the effective clustering of diabetes data. In this approach, we utilize the concept of quantum computing inspired by the aforementioned idea of Han and Kim [20] to evolve the different values of $m$ and $C$ in each generation $g$, so that we can find the global optimal value of these parameters from a large quantum search space. In EQIE-FCM firstly, the fuzziness parameter $m$ in generation $g$ is represented in terms of only one quantum bit as $M'_g$ which is defined as follows:

$$M'_g = Q^m_g \quad (8)$$

Where, $Q^m_g$ consist of two-qubits denoted by $q_p$ where $p = 1, 2$ and represented as $Q^m_g = [q^0_{1m} | q^2_{2m}]$ or $Q^m_g = [\alpha^2_{1m} | \alpha^2_{2m}]$. The reason behind representing $Q^m_g$ in terms of two-qubits because the best value of $m$ has to find within the range of $[1.5, 2.5]$ as suggested by Pal and Bezdek [6], which is the single dimension real value that can be effectively searched from the four subspaces as presented in Section III.

For each value of $m$ represented in terms of quantum bits in generation $g$, the number of clusters $C$ is initialized in the range of $[c_{\min}, c_{\max}]$. In general, for the initialized value of $C$ in generation $g$, the set of cluster centroid is represented as:

$$V^g_C = [(V_{11}^g)^T, (V_{21}^g)^T, \ldots, (V_{i1}^g)^T] \quad (9)$$

Where, $V^g_C$ consist of $t$ number of cluster centroids such that $t = 1, 2, \ldots, C$ and each cluster centroid $(V_{ji}^g)^T$ is represented as follows:

$$V^g_C = [(V_{11}^g)^T, (V_{21}^g)^T, \ldots, (V_{d1}^g)^T] \in R^d \quad (10)$$

where, $(V_{ji}^g)^T$ represents the $j^{th}$ dimension of $i^{th}$ cluster centroid such that $j = 1, 2, \ldots, d$ and $C$ is the number of clusters. For each value of $C$, the set of cluster centroids $V^g_C$ is represented in terms of quantum bits. As stated above and given in Eq (8), the fuzziness parameter $m$ will contain the single dimension real value, therefore only single quantum bit is enough to represent the fuzziness parameter $m$ in a generation $g$. But for each cluster number $C$, the set of cluster centroids $V^g_C$ consist of $d$-dimensions. Thus, multiple quantum bits are required to represent each cluster centroid. However, the $j^{th}$ dimension of $i^{th}$ cluster centroid in generation $g$ is represented in terms of a single quantum bit which is given as follows:

$$(V_{ji}^g)^T = (Q_{ji})^g_C \quad (11)$$

Where, $(Q_{ji})^g_C$ will contain only two-qubits denoted by $q_p$ where $p = 1, 2$ which is represented as $(Q_{ji})^g_C = \left\langle \alpha_{ji} |\alpha_{ji} \right\rangle \beta_{ji} |\beta_{ji} \rangle \quad (12)$
\[ [(q_{ji})^g_C] = [(q_{ji})^g_{2C}] \text{ or } (Q_{ji})^g_C = [(\alpha_{ji})^g_C] = [(\alpha_{ji})^g_{2C}] \text{.} \]

Each cluster centroid in \( j \)-th dimension will represent the real value which can be sufficiently found from the four subspaces as presented in Section III.

It is important to notice that, the proposed algorithm is executed on the classical computer. Therefore, it is required that the quantum value of fuzziness parameter \( M_g \) and the single dimension of the cluster centroid \( (V_{ji})^g_C \) has to be converted into real coded value. This conversion is done with the help of the transformation process. In general, we are presenting the transformation process showing the conversion of single quantum bit \( Q^g \) into real coded value \( (Q')^g \) which is also applicable for the conversion of a quantum value of fuzziness parameter \( M_g \) and the single dimension of the cluster centroid \( (V_{ji})^g_C \). As discussed earlier, in the proposed algorithm we are using only two-qubits denoted by \( q^g_p \) to represent a quantum bit \( Q^g \) where \( p = 1, 2 \). For the conversion of quantum value obtained in generation \( g \) into real coded value, the transformation process starts with the selection of random number vector \( R^g \), where \( R^g = [r^g_1, r^g_2] \) corresponding to the quantum vector \( Q^g = [q^g_1, q^g_2] \) or \( Q^g = [\alpha^g_1, \alpha^g_2] \). Then, further mapping is done by using binary vector \( S^g \) where \( S^g = [s^g_0, s^g_1] \) and the Gaussian random number generator with mean value \( \mu^g_p \) and variance \( \sigma^g_p \), which is represented as \( grg(\mu^g_p, \sigma^g_p) \).

Using the random number vector and the quantum vector, the binary vector \( S^g \) is generated as follows:

\[
\text{if } (r^g_p \leq (\alpha^g_p)^2) \text{ then } s^g_p = 1 \text{ else } s^g_p = 0.
\]

Now with the help of binary vector and the Gaussian random number generator, the real coded value is selected using formula \( \text{bin2dec}(S^g) + 1 \). As presented in Eq 2, the qubit \( (q_p) \) is consisted of two components \( \alpha_p \) and \( \beta_p \). Generally for the processing of qubits, only \( \alpha_p \) is considered because the value of second component \( \beta_p \) will be \( \sqrt{1 - \alpha_p^2} \) [20].

The transformation process shows the conversion of a single quantum bit \( (Q')^g \) into the real coded value is presented in terms of pseudo code as follows:

\[
\text{Transformation process}()
\]

\[
\begin{align*}
\text{begin} \\
\text{Step-1: Initialize quantum vector } Q, \text{ random number vector } R \text{ and link } = 0. \\
\text{for } p := 1 \text{ to } 2 \text{ step } 1 \text{ do} \\
\quad Q^g = \alpha^g_p; \quad 0 \leq \alpha^g_p \leq 1 \\
\quad r^g_p = \text{rand}; \\
\text{end for} \\
\text{Step-2: for } p := 1 \text{ to } 2 \text{ step } 1 \text{ do} \\
\quad \text{if } (r^g_p \leq (\alpha^g_p)^2) \text{ then } s^g_p = 1; \text{ else } s^g_p = 0; \\
\text{end if} \\
\text{end for} \\
\text{Step-3: link } = \text{bin2dec}(S^g) + 1 \\
\text{if } \text{link } = 0 \\
\quad p = \text{link}; \\
\quad (Q')^g = \text{grg}(\mu^g_p, \sigma^g_p); \\
\text{end if} \\
\text{end return } (Q')^g
\end{align*}
\]

Once, the transformation process is completed the real coded value for the fuzziness factor \( m_g \) is represented as follows:

\[
m_g = (Q')^g
\]

Similarly, the real coded value of cluster centroid is represented as follows:

\[
(Q'_{ji})^g_C = (Q')^g
\]

such that

\[
(Q'_{ji})^g_C = [(v'_{1i})^g_C, (v'_{2i})^g_C, ..., (v'_{di})^g_C]^T \in \mathbb{R}^d
\]

\[
(v')^g_C = [(v'_1)^g_C, (v'_2)^g_C, ..., (v'_d)^g_C]
\]

Where, \((v'_{ji})^g_C\) is the real coded value of the \( j \)-th dimension of \( i \)-th cluster centroid and \((v'_j)^g_d\) denote the real coded value of set of cluster centroids such that \( j = 1, 2, ..., d, t = 1, 2, ... C \) and \( C \) is the number of clusters.

In EQIE-FCM algorithm, we evolve the different value of fuzziness parameter and the cluster centroids in each generation by utilizing the quantum rotational gate [21]. It is an important parameter in quantum inspired approaches and used to update the quantum bits of fuzziness parameter and the cluster centroids in each generation. The new qubit is generated using quantum rotational gate and previous value of a qubit which is defined as follows:

\[
\alpha_p^{g+1} = [\alpha_p^g \cos \Delta \theta - \sqrt{1 - (\alpha_p^g)^2} \sin \Delta \theta]
\]

Where, rotational angle \( \Delta \theta \) will provide a proper angle to rotate the quantum bit so that we can get the appropriate value of a new quantum bit. The appropriate angle is selected on the basis of conditions presented in Table II. Furthermore, the value of \( \Delta \theta \) must be selected in such a way so that it can cover a maximum number of values of \( \alpha_p^g \) in the range of \( (0, 1) \) with a minimum number of iterations. Hence, according to Han and Kim [21], \( \Delta \theta \) must be initialized between \( [0.01 \times \pi, 0.05 \times \pi] \).

From preventing the quantum bit \( \alpha_p^g \) from attaining values 0 or 1, following constraints are applied.

\[
\alpha_p^g = \begin{cases} 
\sqrt{\epsilon}, & \text{if } \alpha_p^g < \sqrt{\epsilon} \\
\alpha_p^g, & \text{if } \sqrt{\epsilon} \leq \alpha_p^g \leq \sqrt{1 - \epsilon} \\
\sqrt{1 - \epsilon}, & \text{if } \alpha_p^g > \sqrt{1 - \epsilon}
\end{cases}
\]

Where, the limiting parameter \( \epsilon \) is assigned a very small value (approximately approaching to zero), so that it can cover maximum value in the range of \( (0, 1) \).

\[
\begin{array}{cccc}
\text{TABLE II. PARAMETERS FOR QUBITS updATION.} \\
\hline
s_p^g & s_p^{\text{global}} & F_{\text{UBG2}}(m_{gbest}, C_{best}) > F_{\text{UBG2}}(m_g, C) \text{ or } F_{\text{LBG2}}(m_g, C) > F_{\text{LBG2}}^\prime \text{ at } F_{\text{LBG2}}^\prime > F_{\text{LBG2}}^g \text{ } & \Delta \theta \text{ } \\
\hline
0 & 0 & \text{false} & 0 \text{ } \\
1 & 0 & \text{true} & 0 \text{ } \\
0 & 1 & \text{false} & -0.03 \times \Pi \text{ } \\
1 & 0 & \text{true} & 0 \text{ } \\
1 & 1 & \text{false} & 0 \text{ } \\
0 & 0 & \text{true} & 0 \text{ } \\
1 & 1 & \text{true} & 0 \text{ } \\
0 & 0 & \text{true} & 0 \text{ } \\
\hline
\end{array}
\]

775
The proposed algorithm is executed for \( g_{\text{max}} \) number of generations to find the global optimal value of fuzziness parameter and cluster centroid. The \( g_{\text{max}} \) is set as the stopping criteria for EQIE-FCM algorithm because if it is executed for more than \( g_{\text{max}} \) generations, then it will generate the similar values of these parameters which result in computational overhead. The step-wise procedure of proposed EQIE-FCM algorithm with the above stated parameters is summarized as follows:

**Algorithm 1. EQIE-FCM algorithm**

**Input:** \( X = [x_1, x_2, \ldots, x_N] \); The best location of set of cluster centroids \( v_{\text{best}} \) is initialized as \( \phi \). Initialize the local best fitness function \( F_C^{\text{gbest}} \) and global best fitness function \( F_{C,\text{best}}(m_{\text{best}}, C_{\text{best}}) \) to \( \infty \).

**Process:**

1: The current generation \( g \) is initialized as 1 and set the maximum number of generation \( g_{\text{max}} \) to 100.

2: while \( g \leq g_{\text{max}} \) do

   (A) The fuzziness parameter \( (m) \) for generation \( (g) \) is initialized in terms of quantum bits using Eq (8).

   (B) **Call transformation process** \( (M_g') \): Obtain the real coded value \( m_g \) corresponding to the quantum value \( M_g' \) using transformation process and Eq (12).

   (C) Initialize the parameters like termination criteria \( T \), number of clusters \( C \), \( \Delta \theta \) and \( \epsilon \) using Table III.

   (D) for \( C := c_{\text{min}} \) to \( c_{\text{max}} \) step 1 do

      (I) Initialize criteria function \( J_m((v^{i})_C^g : X, m_g, C, U^g) \) = \( \infty \) and cluster centroids \( (V_j^g)_C \) in terms of quantum bits using Eq (11).

      (II) **Call transformation process** \( (V_j')_C^g \): Obtain the real coded value \( (v^{i})_C^g \) corresponding to the quantum value \( (V_j')_C \) using transformation process and Eq (13).

   (III) repeat

      (a) Compute the fuzzy partition matrix \( U^g = [\mu^g_{il}] \) for \( 1 \leq i \leq C \) and \( 1 \leq l \leq N \).

      \[
      \mu^g_{il} = \frac{\parallel x_l - (v^{i})_C^g \parallel^{-2/m_g}}{\sum_{i=1}^{C} \parallel x_l - (v^{i})_C^g \parallel^{-2/m_g}} \tag{18}
      \]

      (b) Check the fuzzy partition matrix \( U^g \) obtained in Eq (18) satisfy the condition stated below:

      \[
      \sum_{i=1}^{C} \mu^g_{il} = 1 \tag{19}
      \]

      (c) Update the cluster centroids \( (v^{i})_C^g \) for \( 1 \leq i \leq C \).

      \[
      (v^{i})_C^g = \sum_{i=1}^{C} \frac{[(\mu^g_{il})^{m_g}] x_l}{\sum_{i=1}^{C} [(\mu^g_{il})^{m_g}]} \tag{20}
      \]

      (d) Compute the criteria function \( J_m((v^{i})_C^g : X, m_g, C, U^g) \) to evaluate the fitness of obtained fuzzy partition.

      \[
      J_m((v^{i})_C^g : X, m_g, C, U^g) = \sum_{l=1}^{N} \sum_{i=1}^{C} (\mu^g_{il})^{m_g} \parallel x_l - (v^{i})_C^g \parallel^2 \tag{21}
      \]

   until \( (J_m((v^{i})_C^g : X, m_g, C, U^g) \geq T) \)

   end for

   (E) Compute the \( V I_{DSO} \) index [7] which is used as the objective function \( V I_{DSO}(C, U_g) \) in this algorithm to evaluate the fitness of obtained partitions for all the values of \( C \) corresponding to \( m_g \).

   (F) Compute the summation of \( V I_{DSO}(C, U_g) \) as follows:

   \[
   VI_{DSO}^{\text{sum}}(C, U_g) = \sum_{C=c_{\text{min}}}^{c_{\text{max}}} V I_{DSO}(C, U_g) \tag{22}
   \]

   for \( C := c_{\text{min}} \) to \( c_{\text{max}} \) step 1 do

      i) Compute the normalized value of \( V I_{DSO}(C, U_g) \) corresponding to all the values of \( C \).

      \[
      V I_{DSO}^{\text{Normalized}}(C, U_g) = \frac{V I_{DSO}(C, U_g)}{VI_{DSO}^{\text{sum}}(C, U_g)} \tag{23}
      \]

      ii) Store the fitness of fuzzy partition corresponding to each cluster number \( C \) in \( F_C^g \).

      \[
      F_C^g = V I_{DSO}^{\text{Normalized}}(C, U_g) \tag{24}
      \]

   iii) if \( F_C^g \leq F_C^{\text{gbest}} \) then

      \[
      F_C^{\text{gbest}} = F_C^g \quad v_{\text{best}} = (v^{i})_C^g \tag{16}
      \]

      Update the quantum bits of \( (V_j')_C \) by using Table II and Eqs (16) and (17).

   else

      Update the quantum bits of \( (V_j')_C \) by using Table II and Eqs (16) and (17).

   end if

   end for

   (H) Compute local best fitness \( F_{L,\text{best}}^g (m_g, C) \) to determine the best fitness value in generation \( (g) \) as follows:

   \[
   F_{L,\text{best}}^g (m_g, C) = \min_{c_{\text{min}} \leq C \leq c_{\text{max}}} [VI_{DSO}^{\text{Normalized}}(C, U_g)] \tag{25}
   \]

   (I) Compute the global best fitness denoted by \( F_{\text{best}}(m_{\text{best}}, C_{\text{best}}) \) to identify the best value of fuzziness factor and the number of clusters from the overall generations as follows:

   \[
   F_{\text{best}}(m_{\text{best}}, C_{\text{best}}) = \min (F_{\text{gbest}}(m_{\text{best}}, C_{\text{best}}), F_{L,\text{best}}^g (m_g, C)) \tag{26}
   \]

   (J) Update the quantum bits of \( (M_g') \) by using Table II and Eqs (16) and (17).

3: Update \( g = g + 1 \).

4: end while

5: return \( m_{\text{best}}, C_{\text{best}} \) and best location of set of initial cluster centroids \( v_{\text{best}} \).

6: End

**V. EXPERIMENTS**

**A. Experimental Setup and Parameters Specification**

The proposed EQIE-FCM algorithm is implemented in MATLAB computing environment and executed on MATLAB.
version R2014a. The experimentation is done on an Intel(R) Xeon(R) E5-1607 Workstation PC of 3.0 GHz with 64 GB of RAM and running on the Windows 7 Professional operating system. The performance of EQIE-FCM algorithm is compared with QIE-FCM algorithm [16]. The parameter settings of these algorithms are given in Table III. In QIE-FCM algorithm, the set of initial cluster centroids for each cluster number \( C \) is initialized randomly. On the contrary, in case of EQIE-FCM algorithm, it is generated using the quantum bits as mentioned in Section IV.

### B. Results and Discussion

In this section, we present the experimental results to judge the superiority of proposed EQIE-FCM algorithm in comparison with QIE-FCM algorithm. The efficacy of EQIE-FCM algorithm is measured based on following parameters:

1) Evaluation of best fitness value and fuzziness parameter for Pima Indians Diabetes data: The best value of fuzziness parameter and the fitness function achieved by EQIE-FCM algorithm in comparison with QIE-FCM algorithm for Pima Indians Diabetes data are presented in Fig. 2. The comparative result is reported on different values of a fuzziness parameter obtained in 100 generations. In both the algorithms, the \( V_{DSO}^{I} \) [7] index is used as the objective function and the fitness functions used in these algorithms are formulated using this objective function as discussed in Section IV and [16]. The fitness functions formulated in these algorithms are used to evaluate the fitness of produced fuzzy partitions. The small value of \( V_{DSO}^{I} \) in turns reflects the small value of fitness function and thus represents the better fuzzy partitions. Fig. 2, show that the minimum value of the fitness function achieved by EQIE-FCM algorithm is 5.0807E-06 at \( m = 1.5154 \) which is comparatively 6.12 times lesser than the fitness value attain by QIE-FCM algorithm at \( m = 1.5154 \). Although, both the algorithms identify the best value of fuzziness parameter \( m \) at 1.5154 but the EQIE-FCM algorithm achieved the optimal value of fitness function in comparison with QIE-FCM algorithm. Hence, the above reported results justify the superiority of EQIE-FCM algorithm over QIE-FCM algorithms in terms of fitness value.

2) Sensitivity analysis of \( m \) over \( C \): As observed by Pal and Bezdek [6] that fuzzy based clustering algorithms achieved the best clustering results if the fuzziness parameter \( m \) is selected within the range of \([1.5, 2.5]\). In addition to this, researchers also pointed out that these algorithms are considered reliable when the number of clusters \( C \) identified by these approaches is insensitive with change in \( m \). Based on the above consideration in Fig. 3a, we have reported the optimal number of clusters identified by EQIE-FCM algorithm on ten different values of \( m \). It is seen that the number of clusters \( C \) identified by EQIE-FCM algorithm is similar to the number of clusters as per the distribution of data shown in Fig. 1. Moreover, IQIE-FCM algorithm always identifies the same number of clusters on different values of \( m \). Thus, it is inferred that EQIE-FCM algorithm is considered reliable because the number of clusters identified by EQIE-FCM algorithm is insensitive with change in \( m \). The similar observation can be drawn from the Fig. 3b corresponding to QIE-FCM algorithm. Even though, both the algorithms are considered reliable in terms of predicting the number of clusters and also the identified number of clusters is insensitive with change in \( m \). Despite, the value of the fitness function achieved by EQIE-FCM algorithm while predicting the optimal number of clusters \( C \) on different values of \( m \) is comparatively much lesser than the fitness value attained by QIE-FCM algorithm.
3) Comparison of initial cluster centroid location: The initial cluster centroids location predicted by EQIE-FCM algorithm in comparison with the randomly chosen location of cluster centroids by QIE-FCM algorithm is presented in Table IV. In this table, the content highlighted in bold represents the location of cluster centroids found by both the algorithms corresponding to the two-dimensional scatter plot of Pima Indians Diabetes data shown in Fig. 1. It is observed that the initial cluster centroid location predicted by EQIE-FCM algorithm is reasonable because each predicted location of the centroid is almost in the center of the cluster shown in Fig. 1. However, the initial cluster centroid location chosen by QIE-FCM algorithm almost collides with each other. Due to the random selection of the initial cluster centroids by QIE-FCM algorithm, the clustering results achieved by this algorithm may trap into the local optima. Conversely, in EQIE-FCM algorithm the cluster centroids are initially represented in terms of quantum bit and after several generations of evolution, the EQIE-FCM algorithm comes out with the best location of initial cluster centroids. As we can see in Table IV, the initial cluster centroid location predicted by EQIE-FCM algorithm is more accurate than the randomly chosen location by QIE-FCM algorithm thus, the clustering results obtained with EQIE-FCM algorithm will guarantee to achieve the global optimal solution.

4) Computational performance comparison between EQIE-FCM algorithm and QIE-FCM algorithm in terms of iterations count per cluster: The number of iterations required to find the stable cluster centroid on each cluster number by EQIE-FCM in comparison with QIE-FCM algorithm in 100 generations with a step size of 20 generations is reported in Fig. 4. The results show that, the proposed EQIE-FCM algorithm always takes the least number of iterations in comparison with QIE-FCM algorithm for finding the stable cluster centroid on each cluster number. The reported results after every 20 generations show that the QIE-FCM algorithm is much more computationally intensive than the EQIE-FCM algorithm. The reason behind the better computational performance of EQIE-FCM algorithm in comparison with QIE-FCM algorithm is that in QIE-FCM algorithm, the locations of initial cluster centroids are decided randomly and if the data points are located far away from the specified location of initial cluster centroid, then the algorithm will converge slowly by taking many iterations to find the stable cluster centroid. However, in case of EQIE-FCM, due to the selection procedure of the initial cluster centroid it takes the least number of iterations in finding the stable cluster centroid and result in the fast convergence of the algorithm.

VI. COMPARISON WITH EVOLUTIONARY FUZZY CLUSTERING ALGORITHMS

In this section, to further investigate the efficacy of proposed EQIE-FCM algorithm, it is compared with three evolutionary fuzzy based clustering algorithms [8], [9], [11]. These algorithms are also tested on Pima Indian Diabetes data and efficacy is judged in terms of two parameters, i.e. number...
of clusters and value of the fitness function. Table V, shows that the proposed approach is found to be significantly better than compared approaches in terms of finding the optimal value of the fitness function and its corresponding number of clusters. The number of clusters identified by the proposed EQIE-FCM algorithm for Pima Indian Diabetes (PID) data is exactly similar to the number of clusters shown according to the distribution of PID data shown in Fig. 1. Moreover, the optimal value of the fitness function achieved by EQIE-FCM algorithm is comparatively much lesser than the fitness value attained by the other compared approaches. Hence, the discussed results quantify the effectiveness of the proposed algorithm over the compared algorithms.

VII. CONCLUSION

In this work, we have proposed An Enhanced Quantum-inspired Fuzzy C-Means (EQIE-FCM) algorithm. This algorithm is proposed for finding the global optimal value of fuzziness parameter \( m \), the number of clusters \( C \) and the initial cluster centroid \( V_C \) which play an important role in fuzzy based iterative algorithms. In EQIE-FCM algorithm, the clustering of data is performed by evolving these parameters in several generations using the quantum computing concept. The larger search space provided by the quantum computing concept enables us to find the global optimal value of these parameters. To investigate the effectiveness of the proposed algorithm, we tested it on the Pima Indian Diabetes dataset. The performance of the proposed algorithm is compared with another Quantum-inspired Fuzzy Clustering and three evolutionary fuzzy clustering algorithms. The proposed algorithm is found to be very effective and converges to the global optimal value of these parameters. Experimental results show that the EQIE-FCM algorithm found the consistent cluster centroids location as compared to the random initial cluster centroids. This verifies the effectiveness of the proposed approach over other comparable approaches.

REFERENCES


TABLE V. PERFORMANCE COMPARISON OF EQIE-FCM ALGORITHM WITH FUZZY BASED EVOLUTIONARY CLUSTERING ALGORITHMS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fitness function</td>
<td>Number of Clusters</td>
<td>Fitness function</td>
<td>Number of Clusters</td>
</tr>
<tr>
<td>Pima Indian Diabetes</td>
<td>5.08E-06</td>
<td>3</td>
<td>0.00458</td>
<td>16</td>
</tr>
</tbody>
</table>