Quantum-inspired eXplainable Artificial Intelligence for early detection of early-stage Rheumatoid Arthritis in Primary Care

Arka Mitra
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur, West Bengal, India

Mojtaba Kolahdoozi
Department of Electrical Engineering
Queen's University, Ontario, Canada

Jose L. Salmeron
Data Science Lab
Universidad Pablo de Olavide, Seville, Spain

Amir Mohammad Navali
Orthopedic Surgery Department
Shohada University Hospital, Tabriz, Iran

Alireza Sadeghpour
Orthopedic Surgery Department
Shohada University Hospital
Tabriz, Iran

Samira Abbasgholizadeh Rahimi*
Department of Family Medicine
McGill University,
MILA-Quebec AI Institute,
Quebec, Canada
*Corresponding Author

***Abstract*—**
Background: Rheumatoid arthritis (RA)—a chronic, inflammatory disease—causes bone as well as joint erosion, and if untreated, it can lead to patients' disabilities. Early detection of RA can have a key role in prognosis of the disease.
Objectives: We aim to develop an eXplainable Decision Support System (XDSS) to assist primary care providers in early detection of patients with RA.
Methods: Based on the Sparse Fuzzy Cognitive Maps and quantum-learning algorithm, we develop our explainable intelligent system—which is available as a web server—to assist in the detection of RA patients at early stages and classify the severity of their disease into six different levels, collaborating with two specialists in rheumatology and orthopedic surgery. We collected anonymous data of real patients from Shohada University Hospital, Tabriz, Iran and the data has been used for model development. We also compare the results of our model with machine learning methods (e.g., linear discriminant analysis, Support Vector Machines, and K-Nearest Neighbours). The weights obtained from our model are saved and are deployed as part of a web app to give risk intensity scores based on the patient information.
Results and Conclusions: Our proposed model not only outperforms other machine learning methods in terms of accuracy but also, in contrast to the others, our model reveals the relation of the features with one another and gave higher explainability. For future studies, we suggest scaling up the developed app and identifying facilitators and barriers of using this app in clinical practice.

***Keywords*—** eXplanaible Artificial Intellitence, Interpretable Machine Learning, Fuzzy Cognitive maps, Rheumatoid arthritis, Particle Swarm Optimization

I. Introduction

A. The importance of being able to diagnose RA in primary care

Rheumatoid arthritis (RA) is an autoimmune, chronic inflammatory disease [1], [2], characterized by persistent synovitis, systemic inflammation, and autoantibodies (particularly to rheumatoid factor and citrullinated peptide) [3]. The incidence of RA ranges between 0.5% to 1%, and is more common among women and older adults [3]. Aside from social burden, RA carries a substantial individual burden, resulting in "musculoskeletal deficits, with attendant decline in physical function, quality of life, and cumulative comorbid risk" [4].
Primary care physicians can contribute to improved outcomes of RA patients [1]. Primary care, is the gateway into the health care system for all needs and problems and all conditions, including uncommon or unusual ones such as RA [5], [6]. Primary care providers are expected to recognize RA patients as early as possible and refer them to a rheumatologist [7]. Early diagnosis of RA, and consequently early treatment, are essential to better management of RA and have the potential to reduce bone tissue loss and increase favorable outcomes, including remission [3], [8], [9]. However, diagnosis of RA is complex and difficult, and in many patients, early diagnosis is not possible given that clinical indicators are not specific to RA. Indeed, in the early stages of the disease, the typical RA patient has "tender and swollen joints of recent onset, morning joint stiffness, and abnormal laboratory tests such as elevated concentrations of C-reactive protein or erythrocyte sedimentation rate" [3] which can be indicative of RA or other types of arthritis (e.g. reactive arthritis, osteoarthritis, psoriatic arthritis, infectious arthritis, or rarer autoimmune conditions like connective tissue diseases)[3].

RA is a problem that affects a lot of people and negatively impacts their quality of life. Early diagnosis could reduce the negative impact which means that primary care practitioners need to have reliable diagnostic tools. Therefore, the goal of this study is to develop an explainable and intelligent decision support system based on specialty care health professionals (i.e. rheumatologists and orthopedic surgeons) knowledge.

B. Previous works on diagnosis of RA

In previous work [10], we designed a RA diagnosis decision support system by training a 10-node fully-connected Fuzzy Cognitive Map (FCM) and using a particle swarm optimization (PSO) algorithm. Morita et al. [11] proposed a finger joint detection method for RA diagnosis using 45 Japanese RA patients x-ray images, and support vector machines (SVM). Singh et al. [12] used human knowledge as rules for fuzzy logic controller (FLC) for diagnosis of RA, and Montejo et al. [13] used optical tomography images, extracted 594 features from the images, and using five different classifiers, classified images of RA patients.

Despite the attempts, some improvements still are needed in this area: (a) The previous works introduced fully connected networks. Those models have a high number of parameters, so it is possible for the model to memorize the different samples that it is trained on. This increases the chance of overfitting due to increase in complexity of the network [14] and decreases the ability of both interpretability and static analysis of the network. (b) Previous works have considered simple objective functions in their classification process, like classification accuracy. The chance of low generalization is high when one is dealing with small datasets, like the datasets used in the above mentioned works. Also, accuracy might not be the best metric when the training data has an imbalance in the number of classes. Therefore, it is important to tackle this problem by defining the right objective function. In order to overcome the above mentioned limitations, in this study we proposed a novel method based on FCM and a quantum learning algorithm [15], to classify the severity of RA data into six different classes in a way to make it more interpretable and generalizable. The outcome of the interest is detection of RA patients at early stages.

II. Background

A. Fuzzy Cognitive Maps

FCMs have been developed by Kosko [16] and are based on cognitive maps theory [17]. Using causal models, they attempt to mimic human experts' cognitive processes in specific domains. FCM uses a number of concepts and the causal relationships existing between the features for modeling a system, which can be represented as a directed graph [18]. A FCM includes N_n concepts whose values can be shown as Eq. 1.

$$C = [C_1 C_2 \cdots C_{N_n^2}] \tag{1}$$

where, C is a state vector and $C_i \epsilon$[0, 1] represents the value of the i^{th} concept. As the value of a concept approaches +1, its associated activation degree increases. The causal relationship of concepts can be stated in terms of a weight matrix, shown in the Eq. 2.

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \tag{2}$$

where $w_{ij} \epsilon$[-1, +1] shows the value of a weight from the i^{th} to the j^{th} concept. When w_{ij} is a positive number, the i^{th} concept has a positive impact on the j^{th} concept. In other words, any increase in the i^{th} concept causes an increase in the j^{th} concept. The i^{th} concept has a negative impact on the j^{th} concept when w_{ij} is a negative number. In the case of w_{ij}=0, there is no causal relationship between the i^{th} and j^{th} concepts [18]. Since causation does not necessarily mean correlation, w_{ij} need not be equal to w_{ji}, that is, the weight matrix does not need to be a symmetric matrix. Fig. 1 shows a simple, 4-node FCM with its associated weight matrix.

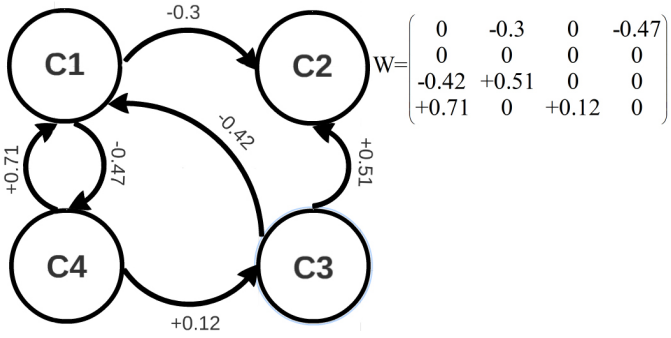

Fig. 1. A 4-node FCM with its weight matrix

The value of the i^{th} node in the (t+1)th iteration can be determined from the weight matrix and the values of the concepts in the t^{th} iteration. By using Eq. 3, one can obtain:

$$C_i(t+1) = \Psi(\sum_{j=1}^{N_n} w_{ij} C_j(t)) \tag{3}$$

where, $\Psi(x)$ is a transfer function, the task of which is to limit the output of the concept values to the desired range. Based on the experiments conducted in [19], sigmoid transfer functions outperform other types of transfer functions; hence, we used this function, stated in the Eq. 4.

$$\Psi(x) = \frac{1}{1 + e^{-\lambda x}} \tag{4}$$

where λ is a free parameter which determines the slope of the function. A typical value of λ is 5 [20]. Consider the Eq. 3 in terms of a matrix multiplication:

$$C(t+1)^T = \Psi(W * C(t)^T) \tag{5}$$

where $A(t)^T$ represents the transpose of matrix A at t^{th} iteration.

The Eq. 5 illustrates that, in every iteration, a FCM calculates the linear combination of row vectors $w_i = [w_{i1} w_{i2} \cdots w_{in}]$,

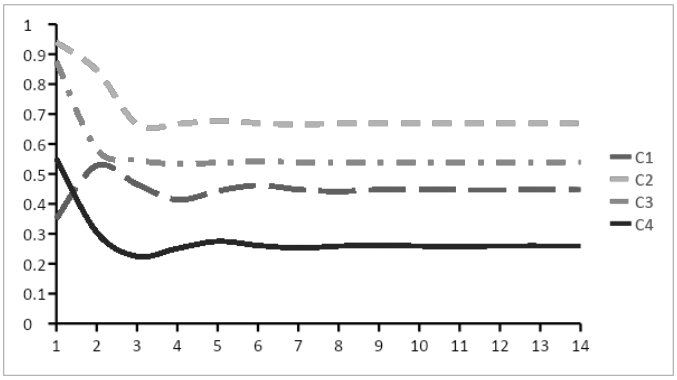

Fig. 2. Fixed point simulation for the FCM shown in Fig. 1

each with the C_i coefficient and does a transformation to keep the values in the desired range. Owing to the use of a continuous transfer function, a FCM simulation can reach one of the following three cases [21]: (1) "Fixed point attractor" where after a limited number of iterations, all concepts converge to a fixed pattern; (2) "Limit cycle", where after a number of iterations, all concepts will fluctuate between a limited number of fixed patterns; and (3) "Chaotic attractor" where concepts will fluctuate between an unlimited number of patterns. Fig. 2 shows a fixed-point attractor simulation for the FCM shown in Fig. 1.

B. Particle Swarm Optimization

Kenndedy and Eberhart [22] introduce Particle Swarm Optimization (PSO) based on behavior observed in nature. It is one of the most popular optimization algorithms and used in various different fields like finance [23], chemistry [24] and medicine [25]. PSO is a population-based search algorithm where the particles comprising the population move in the multi-dimensional space to find the optimal position that optimizes an objective function. Based on the values returned by the objective function at each iteration, the gbest is the position which returns the global best value over all iterations and $pbest_i$ is the position having the best value of the i^{th} particle over all iterations.

The i^{th} position in a d-dimensional search space, denoted by $x_i = [x_i^1, x_i^2, \cdots x_i^d]$, move towards a position in between the gbest and $pbest_i$, guided by velocity v_i which is also a d-dimensional vector. The whole update equations are given in Eqs. 6, 7.

$$x_i(t+1) = v_i(t+1) + x_i(t) \tag{6}$$

$$v_i(t+1) = \omega v_i(t) + c_1 r_1(pbest_i - x_i(t)) + c_2 r_2(gbest - x_i(t)) \tag{7}$$

where ω is a number chosen in the range of [0.1,0.5] and c_1, c_2 are two numbers in the range of [1.5,2]. The values are chosen such that there is a trade-off between exploration and exploitation in the PSO algorithm. More exploration causes the particles to not converge to an optima. While having a lot of exploitation would make the particles get stuck in a local optima, as they are not able to explore most of the search space.

C. The QFCM Algorithm

Fuzzy cognitive maps can be analyzed in two different ways: dynamic and static analysis. In dynamic analysis, values that are obtained from a FCM simulation, and the discrepancies between them and the test pattern are important. In static analysis, the weights, or lack thereof, are important. Non-zero weights in FCM, in contrast to conventional neural networks like multilayer perceptron (MLP), represent a causal relationship between concepts.

Designing algorithms which can form a FCM with both dynamic and static analyses abilities is not an easy task and even conventional algorithms like Non-linear Hebbian Learning (NHL) [26] are not able to do so. Recently, we proposed QFCM algorithm [15] to tackle this problem. It outperformed some other newly developed algorithms like dMAGA [27]. The foundation of the QFCM algorithm is that it models the existence of a weight as a Q-bit, which is the smallest unit of information in the quantum evolutionary algorithm (QEA) [28], and models the values of weights as particles, which are the unit of information in PSO algorithm. Eq. 8 shows a simple Q-bit.

$$Q_i = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \tag{8}$$

In Eq. 8, $|\alpha_i|^2$ and $|\beta_i|^2$ denote the probability that Q_i is found in existence (i.e. one), and inexistence (i.e. zero) states respectively. We combine the quantum evolutionary algorithm (QEA) and particle swarm optimization (PSO) algorithm such that the FCMs, trained by QFCM, not only contain the causal relationship between the components but can also be analyzed dynamically or statically. One of the limitations of the QFCM is that it was developed for time series predictions. It is therefore not currently appropriate for classification problems. In this study, we overcame this limitation.

III. Materials and Methods

Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines are used in this article to report our methods development and validation. The TRIPOD guideline has been developed to support authors in writing reports, aid editors and peer reviewers in reviewing manuscripts submitted for publication and help readers in critically appraising published reports [2]. See supplementary for the TRIPOD checklist.

A. Dataset

To develop our eXplainable decision support system (XDSS), we use a dataset with the information of 13 anonymous patients with RA who were randomly chosen from Shohada University Hospital in 2016 (Table I and Table II). Table. I shows the features that are used in the study along with the justification for their use. Table. II shows some

samples from the dataset and their associated severity or class label. All adult patients, diagnosed with RA were eligible for inclusion in the study. This dataset has been used for training and validating. A subset of this dataset had been used for regression [10]. As with all artificial intelligence(AI) and machine learning(ML) empowered systems, the output of our XDSS is highly related to the data with which it has been developed (input data). Given the complex and ambiguous nature of patient data, including clinical judgements, healthcare professionals may find it easier to express these data using linguistic variables rather than numerical ones [29]. In AI, fuzzy logic can help deal with these ambiguous, subjective, and imprecise judgments. Therefore, with the physicians, we chose six fuzzy variables with Gaussian membership functions (Extremely Severe, Very Severe, Severe, Minor, Very Minor, Extremely Minor) to describe the RA diagnostic criteria. The criteria and justifications for their selection are provided in Table I. For further discussion regarding the selection of these criteria, refer to [10].

Selected Criteria	Justification
C1: Rest pain	Pain is one the most common symptoms in patients with RA. While it is assumed to be interlinked with inflammation, in many cases, despite controlling the inflammation, pain persists [30], [31].
C2: Morning stiffness	This symptom is common among patients with RA. Clinical trials have shown that the duration of this symptom is associated with reduced quality of life [32].
C3: Symmetry of joint infection	Symmetrical joint involvement is a hallmark of RA. Patients usually have several infections in their joints [33].
C4: Redness	Due to inflammation, joints may become red and warm in comparison with the surrounding tissue [33].
C5: Body pain	Patients with RA usually experience moderate and persistent pain in their body [34].
C6: Swelling	One symptom of RA, synovitis, can cause swelling in the joints [35].
C7: Positive Rheumatoid factor (RF) test	This test determines the amount of RF in one's blood. RFs, produced by immune system, are a kind of proteins which are able to destroy healthy tissue. In 70-80% of RA patients test positively for RF. This test has a specificity of 86% [33].
C8: Elevated Erythrocyte sedimentation rate (ESR):	It is a test which is able to determine the severity of inflammation inside a body. It measures the pace at which erythrocytes falls. Patients with RA usually have elevated ESR, owing to hypergammaglobulinemia [33], [36].
C9: Positive Anti-cyclic citrullinated peptide antibody test (Anti-CCP)	57% to 66% of RA patients have a positive-anti-CCP. Positive-anti-CCP patients usually have more severe RA with poor prognosis [33].

TABLE I
CRITERIA FOR DIAGNOSIS OF PATIENTS WITH RA AND THEIR EXPLANATIONS

In addition, based on health professionals' opinions, we assigned six different severity levels to the patients with RA so that they can also help with a more subjective understanding. The levels for each of the conditions for each of the patients is taken and there were no missing data in our dataset. Some of the selected data from the initial data set from the hospital is shown in Table II. The C_i refers to the C_i criteria which is defined in Table. I.

Algorithm 1 The QFCM algorithm, modified for classification problems

```
initialization()
for i = 1...MaxIter do
    for Q = [Q_1 Q_2 ... Q_{n^2}] do
        observe Q to produce a sparse network.
        update velocity and position of the particles.
        mutate particles.
        repair particles.
        classify the RA patients' data by using output concept's value and output fuzzy sets.
        calculate the value of the objective function
        update best local and best global particles
    end for
    update all Qs with H_ε gate.
    update the best quantum candidate.
    if migration period reached then
        perform local as well as global migration.
    end if
end for
```

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	Severity (class label)
1	0.85	0.7	0.5	0.3	0.5	0.7	0.7	0.7	0.7	Extremely severe (5)
2	1.0	0.7	0.5	0.3	0.5	0.7	0.7	0.7	0.7	Extremely severe (5)
3	0.5	0.7	0.5	0.3	0.5	0.3	0.3	0.5	0.5	Very severe (4)
4	0.7	0.5	0.5	0.3	0.5	0.3	0.7	0.7	0.3	Very severe (4)
					. .					
10	0.15	0.15	0.15	0.3	0.15	0.5	0.3	0.5	0.3	Very minor (1)
11	0.0	0.15	0.15	0.0	0.15	0.15	0.15	0.5	0.15	Very minor (1)
12	0.15	0.0	0.15	0.0	0.15	0.15	0.0	0.3	0.15	Extremely minor (0)
13	0.0	0.0	0.15	0.0	0.15	0.15	0.0	0.3	0.15	Extremely minor (0)

TABLE II
SOME OF THE DATASET USED IN THIS STUDY

B. Proposed Method

Our proposed method includes the training of a FCM with our QFCM algorithm [15] modified for classification problems and with a new objective function. The modified version of QFCM algorithm is a supervised learning methodology, that is presented in in algorithm 1.

In the initialization phase, all the Q-bits within a quantum population consist of the training set are initialized with a value of $\frac{1}{\sqrt{2}}$ so that the probability of existence and inexistance of the links becomes equal, ie, $\alpha_i = \beta_i = \frac{1}{\sqrt{2}}$ for all values of i. The positions and velocities of particles, representing the

numerical values of weights, are initialized with a random number ranging between [-1, -0.05) and (+0.05, +1] and 0 respectively. The range of [-0.05, +0.05] is omitted because it cannot represent a causal relationship in a FCM [36]. In the observation process, either 1 (i.e., existence of a link) or 0 (i.e., inexistence of a link) is assigned to the Q-bits, based on the Eq. 9.

$$Bit(Q_i) = \begin{cases} 1, & \text{if } r > |\alpha_i|^2 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

In Eq. 9, r_i is a random number in the range of [0, 1] with uniform distribution. In the next step, the positions and velocities of the particles are updated according to the Eq. 10 and Eq. 11, which are proposed in [37] as the modified version of the PSO algorithm.

$$p_i(t+1) = p_i(t) + v_i(t+1) \tag{10}$$

$$\begin{aligned} v_i(t+1) = \omega v_i(t) + c_i r_i(lbest_i(t) - p_i(t)) \\ + c_2 r_2(gbest(t) - p_i(t)) \end{aligned} \tag{11}$$

where $p_i(t)$ and $v_i(t)$ represent the position and velocity of the i^{th} particle at t^{th} iteration. ω, c1, and c2 are three random numbers in the ranges of [0.1, 0.5], [1.5, 2], and [1.5, 2], respectively. "$lbest_i$" and "gbest" show the best positions of the i^{th} and of all particles, respectively. In step 6 of the QFCM algorithm, mutation occurs: elements from the latter half of each particle are sampled with a probability of μ, and replaced with a random number in the range of [-1,1]. In the repair step (i.e., step 7), the values of all particles are confined to the range [-1, +1] using Eq. 12. It is worth noting that if p_i is in the range (+1,+∞) or in the range (-∞,-1), the velocity of i^{th} particle is multiplied by -1 to reverse the search focus direction given that saturation has occurred in the initial direction. This ensures that the search algorithm does not explore areas that are outside the search space.

$$repair(p_i) = \begin{cases} 0, & \text{if } p_i \epsilon [-0.05, +0.05] \\ +1, & \text{if } p_i \epsilon (+1, +\inf) \\ -1, & \text{if } p_i \epsilon (-\inf, -1) \\ p_i, & \text{otherwise} \end{cases} \tag{12}$$

In the classification step 8, all the trained samples are assigned to one of the six classes, illustrating the severity of RA. To this end, the value of the FCM's output concept is calculated for a given sample, as is the membership degree of this value in each of the six fuzzy sets (Fig. 3). A sample is assigned to a class if its membership degree in this class is higher than that in the other classes. Centers and widths of the membership functions are design parameters.

After the classification step, the output of objective function, proposed in this article in the context of FCMs, is calculated by Eq. 13.

$$F(w) = \frac{\#misclassified}{\#samples} + \sum_{i=1}^{samples} \frac{(x_i - b_i^1)^2 + (x_i - b_i^2)^2}{NF} \tag{13}$$

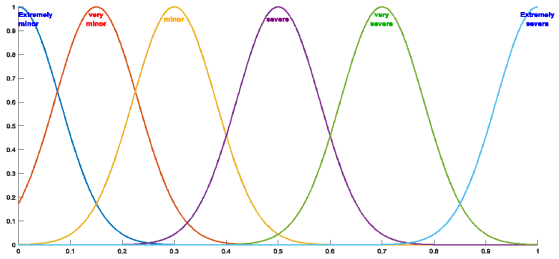

Fig. 3. Membership functions associated with RA severity levels

In Eq. 13, "#misclassified" is the total number of misclassified samples, "#samples" is the total number of samples in the training set, x_i is the value of the FCM output concept for the i^{th} sample in the training set, b_i^1 and b_i^2 are the two borders (i.e., intersection of fuzzy membership functions) nearest to x_i, and NF is the normalization factor that is defined in Eq. 14. NF is defined in order to limit the second term of the objective function to the range of [0, 1].

$$NF = \#Samples \times Lb \tag{14}$$

In Eq. 14, Lb represents the length of the two farthest successive borders. As indicated by Eq. 13, the objective function is designed in such a way that, apart from the classification accuracy, it considers the distance of the training samples from the borders. The global minimum of the second term of the objective function occurs when the FCM maps all the training samples, exactly, to the centers of the successive borders, placing them thus at the furthest possible distance from the borders. Therefore, according to the theory presented in the SVMs [38], the probability of better generalization ability increases. In step 10 of the QFCM algorithm, the best local and best global particles within the quantum population are saved. Subsequently, in step 12, the Q-bits are updated using H gate [39], which is defined in Eq. 15.

$$H_\epsilon(Q_i) = \begin{cases} \begin{bmatrix} \sqrt{\epsilon} \\ \sqrt{1-\epsilon} \end{bmatrix}, & |\alpha|^2 \leq \epsilon \\ Rotate(Q_i), & \epsilon < |\alpha|^2 \leq 1-\epsilon \\ \begin{bmatrix} \sqrt{1-\epsilon} \\ \sqrt{\epsilon} \end{bmatrix}, & 1-\epsilon \leq |\alpha|^2 \end{cases} \tag{15}$$

In Eq. 15, Rotate(Q_i) indicates the rotation of the Q-bit by degrees, and the amount of rotation is a design parameter with the typical value of 0.01. In step 15, local and global migration is done as a mechanism for avoiding local optima. In this regard, values of the best quantum candidate are copied to other candidates locally or globally.

Fengmao et al. [40] showed that after several iterations, the Q-bit converges to either condition 1 or condition 3 of Eq. 15. Kolahdoozi et al. [15] proved that after convergence, it is difficult to escape from the optima it has converged into. Since, the work is an extension to work for classification, the same reason applies and after several iterations, there is very

low probability to escape from the local optima.

The new objective function defined in Eq. 13 considers the predicted labels and the true data to assign values to each position. The modified QFCM algorithm is a supervised learning algorithm that classifies the severity of RA in the patient. For a new patient data, the attributes of the person is taken and the last attribute is taken to be $\frac{1}{\sqrt{2}}$. The attributes for the next iteration is obtained using Eq. 5. The last attribute of the updated list can be mapped onto the fuzzy membership function shown in Fig. 3 to classify the patient into the different categories.

IV. Experimental Results

In this section, we will first present the results of our evaluation on our proposed method and the results of our comparison of the method with other machine learning methods. Then, we will present the contribution of the each of the diagnostic criteria to the results by illustrating the weight matrix obtained from training a FCM with our proposed method. For demonstrating the robustness of the proposed method against different parameter settings, we set the free parameters as shown in Table 3.

MaxIter	Global Migration Period	Local Migration Period	ϵ	μ
1200	20	10	0.01	0.01

TABLE III
VALUES OF THE FREE PARAMETERS OF THE PROPOSED METHOD

A. Classification Accuracy

We trained a 10-node FCM, with one output concept, by using the data shown in the Table II and the proposed method. The dataset consists of only 13 patients taken from randomly from Shohada University Hospital, thus it is delicate to choose a reliable metric. For evaluating its efficacy, in view of the scarce dataset, we used leave-one-out cross validation method (LOOCV). Table. IV-XII shows the accuracy and confusion matrix obtained. Our modified QFCM algorithm (i.e., proposed method) classified nine of the 13 samples correctly, representing an accuracy rate of 69.23%. Among the four misclassified samples, two belong to class 2, one belongs to class 1, and one belongs to class 4. In addition, based on the obtained confusion matrix, in three of the four misclassified samples, the predicted severity is higher than the actual severity. In other words, although misclassified, underestimation of patients with RA is avoided. In clinical contexts, false negatives are extremely dangerous when compared to false positives. Overestimating makes it a false positive rather than a false negative as the patient now has a higher chance of being asked to see a specialist.

In order to compare our results with other machine learning methods, we trained and evaluated different classifiers—namely linear discriminant analysis (LDA), linear SVM, quadratic SVM, cubic SVM, fine K nearest neighborhood (KNN), and weighted KNN—by LOOCV and using the same dataset (Table II). To check the highest accuracy, we also tried to reduce the number of features and rerun the experiments. Since we are removing the search space, methods like KNN should perform better. However, from domain knowledge, it is seen that the features that have been removed to increase accuracy are quite important in clinical experiments. Table. IV-XII presents the results. The two models with fewer features had been checked to see if reducing the features would improve the accuracy or not. In one case, it does increase the accuracy but in cost of losing important clinical features which absolutely needs to be considered in this clinical context. Among the rest of the classifiers evaluated, LDA performed the best with an accuracy rate of 53.8%, which is 15.4% lower than that of our QFCM (i.e. 69.23%). Moreover, unlike our proposed method, LDA underestimates the severity of RA, which may result in misdiagnosis. Fig. 4 presents a coweb [41] graphical representation of our proposed method and LDA to visually compare the two methods. It illustrates that the area under the curve for LDA is larger than that of QFCM illustrating its lower accuracy.

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	0	1	1	0	0	0
2	0	0	0	0	2	0
3	0	0	0	2	0	0
4	0	0	0	1	2	0
5	0	0	0	0	0	2

TABLE IV
PROPOSED METHOD; ACCURACY: 69.23%

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	1	0	1	0	0	0
2	0	0	0	2	0	0
3	0	0	1	0	1	0
4	0	0	0	0	3	0
5	0	0	0	0	0	2

TABLE V
LINEAR DISCRIMINANT ANALYSIS (LDA); ACCURACY: 53.8%

Actual \ Predicted	0	1	2	3	4	5
0	0	2	0	0	0	0
1	1	0	1	0	0	0
2	0	0	0	1	1	0
3	0	0	1	0	1	0
4	0	0	1	0	2	0
5	0	0	0	0	2	0

TABLE VI
LINEAR SVM; ACCURACY 15.4%

B. Weight matrix of the FCM and its associated interpretability

Using our data set (Table II), we trained a FCM, with the weight matrix shown in Eq. 16. The density of this FCM is 50%, meaning that half of the 100 weights are zero. The first nine columns represent the nine criteria in the order presented

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	0	1	0	0	0	0
2	0	0	0	2	0	0
3	0	0	2	0	0	0
4	0	0	1	1	1	0
5	0	0	0	0	0	2

TABLE VII
QUADRATIC SVM; ACCURACY 46.2%

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	1	0	1	0	0	0
2	0	0	0	2	0	0
3	0	0	2	0	0	0
4	0	0	1	1	1	0
5	0	0	0	0		2

TABLE VIII
CUBIC SVM; ACCURACY 38.5%

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	1	1	0	0	0	0
2	0	0	0	1	1	0
3	0	0	2	0	0	0
4	0	0	1	1	1	0
5	0	0	0	0	0	2

TABLE IX
FINE KNN; ACCURACY 46.2%

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	1	0	1	0	0	0
2	0	0	0	1	1	0
3	0	0	1	0	1	0
4	0	0	1	0	2	0
5	0	0	0	0	0	2

TABLE X
WEIGHTED KNN; ACCURACY 46.2%

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	0	2	0	0	0	0
2	0	0	0	2	0	0
3	0	0	0	1	1	0
4	0	0	0	0	3	0
5	0	0	0	0	0	2

TABLE XI
KNN WITH 4 FEATURES; ACCURACY 76.9%

in Table I. Furthermore, an extra node has been added which is connected to all the other nodes. This 10^{th}node is used to determine the contribution of the other nodes to detect the disease. The 10^{th}column of this matrix elucidates the impact of each of the features on the output concept. None of the weights of associated with RA diagnostic tests (i.e., C7, C8, C9) are 0, demonstrating the importance of these tests relative to the physical symptoms of RA. Among the

Actual \ Predicted	0	1	2	3	4	5
0	2	0	0	0	0	0
1	0	1	1	0	0	0
2	0	0	0	0	2	0
3	0	0	1	0	1	0
4	0	0	0	0	3	0
5	0	0	0	0	0	2

TABLE XII
KNN WITH 3 FEATURES; ACCURACY 61.5%

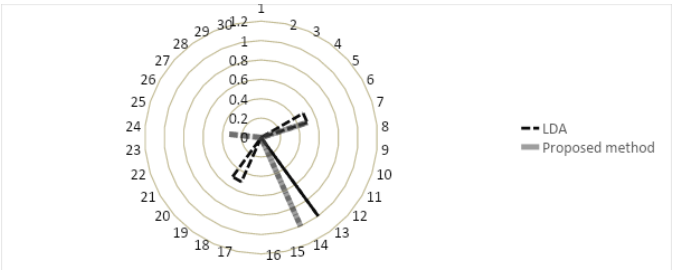

Fig. 4. Cobweb graphical representation of LDA and our proposed method

physical symptoms chosen for the diagnostic criteria, rest pain had the most important contribution to the output, whereas the weights of morning stiffness, redness, and body pain were zero and among lab tests, ESR had a greater impact on the output. Regarding Anti-CCP and RF, our QFCM algorithm assigned a larger weight to Anti-CCP, which indicates that it contributes more to the output than RF, which is compatible with the clinical study conducted on over 1,025 patients [42].

$$W = \begin{bmatrix} -0.906 & 0 & 0 & 0 & 0 & -0.774 & -0.068 & 0 & -0.058 & -0.930 \\ 0 & -0.283 & 0 & 0 & 0 & 0.799 & -0.999 & -0.360 & 0 & 0 \\ -0.707 & 0 & 0 & -0.130 & -0.706 & 0 & 0 & -0.326 & -0.839 & -0.313 \\ 0.137 & 0 & 0.869 & 0.889 & -0.978 & -0.512 & -0.332 & 0 & -0.614 & 0 \\ 0 & -0.738 & 0 & 0 & 0.375 & 0.954 & 0 & 0.749 & 0 & 0 \\ -0.292 & -0.824 & 0.778 & 0 & 0 & 0 & 0 & 0 & 0.852 & 0.647 \\ -0.612 & 0.416 & 0 & 0 & 0 & 0 & 0 & -0.215 & -0.275 & -0.616 \\ 0.869 & -0.937 & 0 & 0 & -0.735 & 0 & 0 & -0.877 & 0 & -0.999 \\ 0 & 0 & 0 & 0.945 & 0.393 & 0.444 & 0 & -0.484 & 0 & 0.623 \\ 1 & 0 & 0 & -0.403 & -0.787 & 0 & 0 & 0 & 0.447 & 0 \end{bmatrix} \quad (16)$$

Using the Eq. 16, the interactions between the criteria can be investigated. Weights with values near to 1 or -1 are indicative of strong relationships. For example, referring to the first column on the left, if we ignore the self-feedback/loop, our results indicate that ESR (i.e., C9) is the criterion most strongly related to rest pain (i.e., C1) and symmetry of joint infection (i.e., C3), or according to the 5^{th}column from the left, body pain and redness (i.e., C5 and C4) are interlinked.

C. Web Based App

Our XDSS is freely available for academic purposes and can be accessed from the github page https://github.com/rahimi-s-lab/RA-paper and is coded in the Hypertext Preprocessor (PHP) language to make it easy to use(Fig. 5). To use it to help identify an RA patient, the input data should be uploaded as a text file with the patient data for each of the nine diagnostic criteria. This would allow large number of patients to be classified at once since the text file accepts multiple patient data. The XDSS will perform all calculations

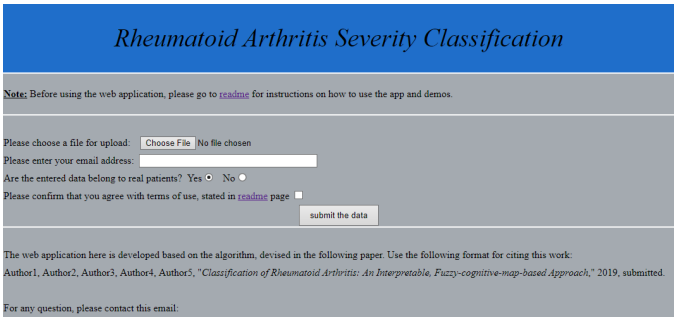

Fig. 5. Demonstration version of our developed web-based eXplainable Decision Support System

and immediately display the patient's severity of RA along with interpretations of the results.

V. Limitations

We acknowledge that the dataset that was obtained for the study was relatively low in sample size. At the same time ,this is a good example on how small dataset could be used in the context with the proposed model and extract the causal relationships between the different parameters as shown in 16. Our domain experts confirmed the results and explored relationships, however, a larger dataset could have helped us to explore a better model.

VI. Conclusion

Primary care professionals are responsible for identifying patients with RA and referring them to a rheumatologist, however, the diagnosis of patients with RA is complex and, in many patients, early diagnosis by their primary care provider is difficult due to the non-specific nature of their symptoms and clinical indicators. The aim of this study was to: 1) contribute to the existing methodology in the field by overcoming the current limitations and 2) develop a web-based eXplainable Decision Support System (XDSS) to assist primary care professionals in early diagnosis of patients with RA. We develope this XDSS based on well-known soft computing method, Fuzzy Cognitive Maps (FCMs), and modified quantum learning algorithm. To develop algorithm for this XDSS, we consult with two health professionals (a rheumatologist and an orthopedic surgeon) and integrate their knowledge into our model and used the data of actual patients with RA obtained from Shohada University Hospital. We evaluate the accuracy of the QFCM and compared its accuracy rate with other machine learning methods. Our proposed hybrid method obtained highest accuracy and other outperformed machine learning methods. Apart from having higher accuracy, one of the strengths of our proposed hybrid method is its interpretability. Going forward, we will work with primary health care providers to further develop our web-based XDSS such that its design is user-centered, perform larger-scale testing, adapt it to other clinical contexts, and include interlinking the knowledge obtained from the interpretability of the network into human knowledge.

Competing Interests

Authors have no competing interests to declare.

Funding source(s) and role(s)

We didn't obtain any funding for this study.

Acknowledgements

The authors would like to thank the physicians who labelled the data. We would also like to thank Shohada University Hospital for providing the dataset. SR is funded by a Research Scholar Junior 1 Career Development Award by the Fonds de Recherche du Québec-Santé (FRQS), and her research program is supported by Natural Sciences and Engineering Research Council (NSERC) Discovery Grant #2020-05246.

References

[1] F. Mizoguchi, K. Slowikowski, K. Wei, J. L. Marshall, D. A. Rao, S. K. Chang, H. N. Nguyen, E. H. Noss, J. D. Turner, B. E. Earp, P. E. Blazar, J. Wright, B. P. Simmons, L. T. Donlin, G. D. Kalliolias, S. M. Goodman, V. P. Bykerk, L. B. Ivashkiv, J. A. Lederer, N. Hacohen, P. A. Nigrovic, A. Filer, C. D. Buckley, S. Raychaudhuri, and M. B. Brenner, "Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis," *Nature Communications*, vol. 9, no. 1, Feb. 2018. [Online]. Available: https://doi.org/10.1038/s41467-018-02892-y

[2] A.-B. G. Blavnsfeldt, A. de Thurah, M. D. Thomsen, S. Tarp, B. Langdahl, and E.-M. Hauge, "The effect of glucocorticoids on bone mineral density in patients with rheumatoid arthritis: A systematic review and meta-analysis of randomized, controlled trials," *Bone*, vol. 114, pp. 172–180, Sep. 2018. [Online]. Available: https://doi.org/10.1016/j.bone.2018.06.008

[3] D. L. Scott, F. Wolfe, and T. W. Huizinga, "Rheumatoid arthritis," *The Lancet*, vol. 376, no. 9746, pp. 1094–1108, Sep. 2010. [Online]. Available: https://doi.org/10.1016/s0140-6736(10)60826-4

[4] G. D. Kitas and S. E. Gabriel, "Cardiovascular disease in rheumatoid arthritis: state of the art and future perspectives," *Annals of the Rheumatic Diseases*, vol. 70, no. 1, pp. 8–14, Nov. 2010. [Online]. Available: https://doi.org/10.1136/ard.2010.142133

[5] J. Smith, "Primary care: balancing health needs, services and technology," *International Journal of Integrated Care*, vol. 1, p. e36, Sep 2001, pMC1484414[pmcid]. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1484414/

[6] L. A. Green, G. E. Fryer, B. P. Yawn, D. Lanier, and S. M. Dovey, "The ecology of medical care revisited," *New England Journal of Medicine*, vol. 344, no. 26, pp. 2021–2025, Jun. 2001. [Online]. Available: https://doi.org/10.1056/nejm200106283442611

[7] D. L. Goldenberg, "The primary care provider's role in diagnosing and treating rheumatoid arthritis," *Practical Pain Management*, vol. 17, no. 5, Sep 2017.

[8] B. Heidari, "Rheumatoid arthritis: Early diagnosis and treatment outcomes," *Caspian journal of internal medicine*, vol. 2, no. 1, pp. 161–170, 2011, pMC3766928[pmcid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24024009

[9] A. Finckh, "Early inflammatory arthritis versus rheumatoid arthritis," *Current Opinion in Rheumatology*, vol. 21, no. 2, pp. 118–123, Mar. 2009. [Online]. Available: https://doi.org/10.1097/bor.0b013e3283235ac4

[10] J. L. Salmeron, S. Rahimi, A. Navali, and A. Sadeghpour, "Medical diagnosis of rheumatoid arthritis using data driven pso-fcm with scarce datasets," *Neurocomputing*, vol. 232, pp. 104–112, 2017.

[11] K. Morita, A. Tashita, M. Nii, and S. Kobashi, "Computer-aided diagnosis system for rheumatoid arthritis using machine learning," in *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*. IEEE, Jul. 2017. [Online]. Available: https://doi.org/10.1109/icmlc.2017.8108947

[12] S. Singh, A. Kumar, K. Panneerselvam, and J. J. Vennila, "Diagnosis of arthritis through fuzzy inference system," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1459–1468, Oct. 2010. [Online]. Available: https://doi.org/10.1007/s10916-010-9606-9

[13] L. D. Montejo, J. Jia, H. K. Kim, U. J. Netz, S. Blaschke, G. A. Muller, and A. H. Hielscher, "Computer-aided diagnosis of rheumatoid arthritis with optical tomography, part 1: feature extraction," *Journal of Biomedical Optics*, vol. 18, no. 7, p. 076001, Jul. 2013. [Online]. Available: https://doi.org/10.1117/1.jbo.18.7.076001

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. USA: Wiley-Interscience, 2000.

[15] M. Kolahdoozi, A. Amirkhani, M. H. Shojaeefard, and A. Abraham, "A novel quantum inspired algorithm for sparse fuzzy cognitive maps learning," *Applied Intelligence*, vol. 49, no. 10, pp. 3652–3667, May 2019. [Online]. Available: https://doi.org/10.1007/s10489-019-01476-7

[16] B. Kosko, "Fuzzy cognitive maps," *International Journal of Man-Machine Studies*, vol. 24, no. 1, pp. 65–75, Jan. 1986. [Online]. Available: https://doi.org/10.1016/s0020-7373(86)80040-2

[17] E. C. Tolman, "Cognitive maps in rats and men." *Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948. [Online]. Available: https://doi.org/10.1037/h0061626

[18] J. L. Salmeron, T. Mansouri, M. R. S. Moghadam, and A. Mardani, "Learning fuzzy cognitive maps with modified asexual reproduction optimisation algorithm," *Knowledge-Based Systems*, vol. 163, pp. 723–735, Jan. 2019. [Online]. Available: https://doi.org/10.1016/j.knosys.2018.09.034

[19] S. Bueno and J. L. Salmeron, "Benchmarking main activation functions in fuzzy cognitive maps," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5221–5229, Apr. 2009. [Online]. Available: https://doi.org/10.1016/j.eswa.2008.06.072

[20] W. Stach, W. Pedrycz, and L. A. Kurgan, "Learning of fuzzy cognitive maps using density estimate," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 900–912, Jun. 2012. [Online]. Available: https://doi.org/10.1109/tsmcb.2011.2182646

[21] J. L. Salmeron, "Fuzzy cognitive maps for artificial emotions forecasting," *Applied Soft Computing*, vol. 12, no. 12, pp. 3704–3710, Dec. 2012. [Online]. Available: https://doi.org/10.1016/j.asoc.2012.01.015

[22] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948 vol.4.

[23] S. Lahmiri, "Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression," *Applied Mathematics and Computation*, vol. 320, pp. 444–451, Mar. 2018. [Online]. Available: https://doi.org/10.1016/j.amc.2017.09.049

[24] G. Jana, A. Mitra, S. Pan, S. Sural, and P. K. Chattaraj, "Modified particle swarm optimization algorithms for the generation of stable structures of carbon clusters, cn (n = 3–6, 10)," *Frontiers in Chemistry*, vol. 7, Jul. 2019. [Online]. Available: https://doi.org/10.3389/fchem.2019.00485

[25] H. Hu, H. Wang, Y. Bai, and M. Liu, "Determination of endometrial carcinoma with gene expression based on optimized elman neural network," *Applied Mathematics and Computation*, vol. 341, pp. 204–214, Jan. 2019. [Online]. Available: https://doi.org/10.1016/j.amc.2018.09.005

[26] E. Papageorgiou, C. Stylios, and P. Groumpos, "Fuzzy cognitive map learning based on nonlinear hebbian rule," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2003, pp. 256–268. [Online]. Available: https://doi.org/10.1007/978-3-540-24581-0_22

[27] J. L. Salmeron, A. Ruiz-Celma, and A. Mena, "Learning FCMs with multi-local and balanced memetic algorithms for forecasting industrial drying processes," *Neurocomputing*, vol. 232, pp. 52–57, Apr. 2017. [Online]. Available: https://doi.org/10.1016/j.neucom.2016.10.070

[28] K.-H. Han and J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, Dec. 2002. [Online]. Available: https://doi.org/10.1109/tevc.2002.804320

[29] S. A. Rahimi, A. Jamshidi, A. Ruiz, and D. Ait-kadi, "A new dynamic integrated framework for surgical patients' prioritization considering risks and uncertainties," *Decision Support Systems*, vol. 88, pp. 112–120, Aug. 2016. [Online]. Available: https://doi.org/10.1016/j.dss.2016.06.003

[30] M. Löfgren, C. H. Opava, I. Demmelmaier, C. Fridén, I. E. Lundberg, B. Nordgren, and E. Kosek, "Pain sensitivity at rest and during muscle contraction in persons with rheumatoid arthritis: a substudy within the physical activity in rheumatoid arthritis 2010 study," *Arthritis Research & Therapy*, vol. 20, no. 1, Mar. 2018. [Online]. Available: https://doi.org/10.1186/s13075-018-1513-3

[31] C. C. Mok, "Morning stiffness in elderly patients with rheumatoid arthritis: What is known about the effect of biological and targeted agents?" *Drugs & Aging*, vol. 35, no. 6, pp. 477–483, Apr. 2018. [Online]. Available: https://doi.org/10.1007/s40266-018-0548-0

[32] S. West, *Rheumatology Secrets E-Book*. Elsevier Health Sciences, 2019.

[33] D. F. McWilliams and D. A. Walsh, "Pain mechanisms in rheumatoid arthritis," *Clin Exp Rheumatol*, vol. 35, no. Suppl 107, pp. 94–101, 2017.

[34] T. Marhadour, S. Jousse-Joulin, G. ChalèS, L. Grange, C. Hacquard, D. Loeuille, J. Sellam, J.-D. Albert, J. Bentin, I. C. Valckenaere, M.-A. d'Agostino, F. Etchepare, P. Gaudin, C. Hudry, M. Dougdos, and A. Saraux, "Reproducibility of joint swelling assessments in long-lasting rheumatoid arthritis: Influence on disease activity score-28 values (SEA-repro study part i)," *The Journal of Rheumatology*, vol. 37, no. 5, pp. 932–937, Apr. 2010. [Online]. Available: https://doi.org/10.3899/jrheum.090879

[35] Z. Isiksacan, O. Erel, and C. Elbuken, "A portable microfluidic system for rapid measurement of the erythrocyte sedimentation rate," *Lab on a Chip*, vol. 16, no. 24, pp. 4682–4690, 2016. [Online]. Available: https://doi.org/10.1039/c6lc01036a

[36] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Genetic learning of fuzzy cognitive maps," *Fuzzy Sets and Systems*, vol. 153, no. 3, pp. 371–401, Aug. 2005. [Online]. Available: https://doi.org/10.1016/j.fss.2005.01.009

[37] M. R. Sierra and C. A. Coello Coello, "Improving pso-based multi-objective optimization using crowding, mutation and ϵ-dominance," in *Evolutionary Multi-Criterion Optimization*, C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 505–519.

[38] L. Wang, Ed., *Support Vector Machines: Theory and Applications*. Springer Berlin Heidelberg, 2005. [Online]. Available: https://doi.org/10.1007/b95439

[39] K.-H. Han and J.-H. Kim, "Quantum-inspired evolutionary algorithms with a new termination criterion, $ h_epsilon $ gate, and two-phase scheme," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 2, pp. 156–169, Apr. 2004. [Online]. Available: https://doi.org/10.1109/tevc.2004.823467

[40] F. Lv, G. Yang, W. Yang, X. Zhang, and K. Li, "The convergence and termination criterion of quantum-inspired evolutionary neural networks," *Neurocomputing*, vol. 238, pp. 157–167, May 2017. [Online]. Available: https://doi.org/10.1016/j.neucom.2017.01.048

[41] B. Diri and S. Albayrak, "Visualization and analysis of classifiers performance in multi-class medical data," *Expert Systems with Applications*, vol. 34, no. 1, pp. 628–634, Jan. 2008. [Online]. Available: https://doi.org/10.1016/j.eswa.2006.10.016

[42] I. G. Silveira, R. W. Burlingame, C. A. von Mühlen, A. L. Bender, and H. L. Staub, "Anti-CCP antibodies have more diagnostic impact than rheumatoid factor (RF) in a population tested for RF," *Clinical Rheumatology*, vol. 26, no. 11, pp. 1883–1889, Apr. 2007. [Online]. Available: https://doi.org/10.1007/s10067-007-0601-6