A New Type of Fuzzy-Rule-Based System With Chaotic Swarm Intelligence for Multiclassification of Pain Perception From fMRI

Ahmed M. Anter, Gan Huang, Linling Li, Li Zhang, Zhen Liang, and Zhiguo Zhang

***Abstract*—Machine learning has been increasingly used in decoding brain states from functional magnetic resonance imaging (fMRI). One important application is to classify the levels of pain perception from patients' fMRI for clinical pain assessment. However, the huge number of fMRI features and the complex relationships between fMRI and pain levels affect the performance of pain classification models heavily. In this article, we introduce a new fuzzy-rule-based hybrid optimization approach for dimension reduction and multiclassification problems using chaotic map, crow search optimization (CSO), and self-organizing fuzzy logic prototype (SOFLP). The approach is named as CCSO–SOFLP. In the proposed approach, chaotic map-based CSO is employed to find the optimal features from ultra-high-dimensional fMRI, and the fuzzy-rule-based SOFLP is employed for multiclassification of pain levels. In this sense, CSO is provided to avoid being stuck in local minima and to increase the computational performance. On the other hand, multilayer SOFLP classifier can continuously learn from new data and identify prototypes from the observed data and use them to build fuzzy rules, to define a suitable local area for each prototype, and to avoid overlapping. The proposed approach is applied on a pain-evoked fMRI data set to classify the levels of pain. Results indicate that the proposed approach can decode levels of pain and identify predictive fMRI patterns with higher accuracy and convergence speed and shorter execution time. Therefore, the new type of fuzzy-rule-based system with chaotic swarm intelligence holds great potential to predict pain perception in clinical uses.**

***Index Terms*—Crow search optimization (CSO), functional magnetic resonance imaging (fMRI) decoding, fuzzy rules, pain prediction, self-organizing fuzzy logic prototype (SOFLP).**

Manuscript received August 18, 2019; revised December 24, 2019 and February 24, 2020; accepted February 25, 2020. Date of publication March 6, 2020; date of current version June 3, 2020. This work was supported in part by National Natural Science Foundation of China under Grant 81871443, Grant 61701316, Grant 81901831, and Grant 61906122, in part by the Science, Technology and Innovation Commission of Shenzhen Municipality Technology Fund under Grant JCYJ20170818093322718, and in part by Shenzhen Peacock Plan under Grant KQTD2016053112051497. *(Corresponding author: Zhiguo Zhang.)*

Ahmed M. Anter is with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China, and also with the Faculty of Computers and Artificial Intelligence, Beni-Suef University, Benisuef 62511, Egypt (e-mail: anter@szu.edu.cn, ahmed_anter@fcis.bsu.edu.eg).

Gan Huang, Linling Li, Li Zhang, and Zhen Liang are with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China (e-mail: huanggan@szu.edu.cn; lilinling@szu.edu.cn; lzhang@szu.edu.cn; janezliang@szu.edu.cn).

Zhiguo Zhang is with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: zgzhang@szu.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2020.2979150

I. Introduction

PAIN is a highly subjective experience, and pain assessment is crucial in clinical practice. Self-report is the gold standard for pain assessment, but in some populations, such as coma patients, children, cognitive impairment patients, and individuals who are unwilling to communicate their feelings of pain, it may not be available or reliable [1]–[3]. Inaccurate pain assessment can lead to insufficient leadership of pain and even mislead diagnosis and treatment [4]. Therefore, the development of a neurophysiology-based pain assessment tool is highly necessary in the fundamental pain research and clinical applications [5]. Recently, the classification of pain level intensity from neuroimaging data, such as functional magnetic resonance imaging (fMRI), has attracted increasing interests and is regarded to have the ability to provide physiological and quantitative pain evaluation tools that complement self-reporting [6]–[9].

However, current fMRI-based pain classification models reported in literatures can hardly be used clinically, because the efficiency of fMRI decoding is hampered by ultra-high dimension and nonlinear information [10], [11]. Therefore, dimension reduction techniques should be applied to extract most predictive patterns or feature sets before prediction. Also, most related pain-prediction studies only deal with binary classification (high pain versus low pain) [10]–[13], but clinical pain assessment requires to classify pain into 11 levels (from 0 to 10). Therefore, multiclassification of pain levels is highly desired.

Generally, there are two main feature reduction approaches for fMRI decoding: filter approach (such as Fisher score, relief, mutual information, and information gain) and wrapper approach (such as sequential backward search, sequential forward search, and sequential random search). Almost all of these algorithms are computationally expensive and could be stuck in local minima [14]. To further improve the efficiency of feature reduction, a global search-based swarm intelligence (SI) algorithm is required [15]–[17].

The crow search optimization (CSO) is one of the SI algorithms which is a population-based metaheuristic optimization algorithm and can be utilized for feature reduction problems [18]. The CSO follows the mechanism of crows' searching to store their food in hiding places and to retrieve the food in times of need. Only with two adjustable parameters [flight length (FL) and awareness probability (AP)], CSO becomes less computationally expensive than many other metaheuristic algorithms,

which need many parameters to adjust and can provide a user-friendly optimization model for different applications. In defined number of iterations, the CSO algorithm tries to find the optimal solution (the best crow position obtained so far) to represent the optimization problem. The main issue with CSO is the low rate of convergence due to the trapping in local optima because there is no clear boundary between exploration and exploitation due to the stochastic nature of CSO [19]. A lot of exploration leads to a lot of new solutions, which are time-consuming, and the optimizer may be stuck in local minima with no convergence. Thus, balancing between exploration (diversification) and exploitation (intensification) during the optimization process is a difficult issue. Chaotic map is proposed in this article to balance between exploration and exploitation rates and to enhance SI efficiency both in terms of local minima prevention and convergence speed [20].

On the other hand, machine learning (ML) has gained popularity in the community of brain science and engineering for decoding stimuli, mental states, behaviors, and other interesting variables from neuroimaging data. Many ML algorithms, however, have some prevalent disadvantages. First, they cannot be incrementally learned, which does not allow to correct errors of the recognition dynamically. Second, they do not coverage well to optimal solution and are easily stuck in local minima. Third, they have a large quantity of parameters and need prior knowledge of the problem, which makes it hard to adjust their performance, such as radius, learning rate, decay rates, and size of the network [21], [22].

During the last four decades, the fuzzy sets and fuzzy-rule-based (FRB) systems have been emerged and are commonly accepted as a dominant mechanism and framework for capturing and representing intelligent systems [23]–[25]. As it has the benefits of high transparency and interpretability of outcomes, FRB systems have become one of the alternative frameworks for ML design. Two famous FRB systems are commonly used—Mamdani and Takagi–Sugeno (TS). Both Mamdani and TS types share the exact same antecedent (if) part and only (although significantly) differ by the consequent (then) part. They are distinguishable by their consequent part, which is crisp for the TS type whereas fuzzy sets based for the Mamdani type [26]–[28].

There are a number of issues with such an FRB approach [29], [30], including the following.

1) The degree of activation of a fuzzy rule is determined as an aggregation of the degrees of memberships of a data sample to each of the fuzzy sets.
2) Membership functions are determined by experts and needs parameterization to determine the center and left/right boundaries or spread.
3) Membership functions are often differ significantly from the real data distribution.
4) Fuzzy clustering memberships ignore or approximate all data forming the cluster.

Motivated by the abovementioned discussions, this study proposes a novel dynamic approach based on chaotic map, binary CSO algorithm, and self-organizing fuzzy logic prototype (SOFLP) algorithm to the complex and nonlinear problem of fMRI-based multiclassification of pain intensity. More specifically, the hybrid chaotic map and binary CSO (CCSO) approach is adopted in this study to select the optimal subset of features from the ultra-high-dimensional fMRI data. The CCSO approach uses the strong ability of the global optimizing of the CSO with a smaller number of iterations, less time consumption, and avoids the sensitivity to local optimization. Furthermore, the chaotic map handles the lack of convergence of the CSO algorithm, which transferred the random variables from Gaussian distribution to chaotic behavior. Moreover, the SOFLP approach is based on AnYa-type FRB, which keeps the advantages of traditional FRB systems and avoids the problems related to membership functions, identification, and update. SOFLP is a nonparametric data-driven, memory efficient system structure that can learn from the newly arrived data samples, adapt with nonstationary environment by updating metaparameters recursively, and change continuously its structure. The proposed CCSO–SOFLP approach is computationally efficient and validated against the ML methods with respect to different measures.

Moreover, this article is aimed to quantitatively analyze the pain prediction from pain-evoked fMRI, and more specifically, to identify cortical regions that are affected by pain stimuli. Using a laser-evoked pain experiment, fMRI data and pain ratings are collected to achieve this aim. To the best of our knowledge, there is no fuzzy logic and metaheuristic optimization work on the topic of pain prediction. The proposed CCSO–SOFLP approach is very useful because it allows the relationships between behavior responses and brain activities to be implicitly modeled. Due to the complicated neurovascular mechanism, a clear hypothesis to model the connection between behavioral reactions and brain activities is often hard to formulate. Therefore, the proposed approach holds great potential in basic research and clinical applications for pain prediction. Overall, the novelty of the proposed approach is threefold.

1) First, a new chaotic CSO (CCSO) method is proposed to select the optimal subset of pain-related features from high-dimensional neuroimaging data. The new CCSO method is a binary version of CSO, and it jointly incorporates a sigmoid transfer function, k-nearest neighbor (KNN), and chaotic map to make the conventional CSO suitable for feature selection. More precisely, the sigmoid transfer function is used to transfer agents from continuous to binary space to fit the feature selection problem. The chaotic sequence map is utilized to replace the random behavior to a random chaotic motion in CSO, and it can achieve balance between the exploration and exploitation rates. The KNN method is used as a cost function and works as a part of the optimizer to assess every crow position in the search space.
2) Second, it may be the first attempt to solve the problem of imbalanced pain-evoked fMRI data. Although it is well known that the levels of pain perception are highly imbalanced in pain-related data samples, there are few studies aiming to solve this problem using oversampling/undersampling techniques. In this article, we apply the synthetic minority oversampling technique (SMOTE) method to generate synthetic data, which can adjust the class distribution of the pain data set.

3) Third, the whole CCSO–SOFLP approach is a new and effective pipeline for fMRI-based pain prediction. In this approach, CCSO is used for feature selection, and SOFLP is used for multiclass classification. The multilayer FRB SOFLP method based on AnYa-type FRB is a recently proposed memory- and computation-efficient classification method. The CCSO–SOFLP approach is free from prior assumptions and problem-specific parameters, while its training process is highly parallelizable.

The rest of this article is structured as follows. Section II introduces preliminary concepts of the CSO algorithm, imbalanced fMRI data, and SOFLP approach. Section III provides the experimental materials and data acquisition and presents the proposed CCSO–SOFLP approach for multiclassification of pain intensity. In Section IV, experimental results are presented. Finally, Section V concludes this article.

II. Background of Methods Involved

In this section, all methods utilized in the proposed approach are introduced. Some traditional methods, such as augmentation of imbalanced fMRI data, some basic theories of the CSO algorithm and chaotic map for tuning CSO parameters, and the SOFLP method based on AnYa-type FRB, are presented.

A. Augmentation of Imbalanced fMRI Data

The issue of imbalanced information appears when class samples have a majority class that outnumbering the other (minority class) samples considerably. Hence, learning from a minority class is hard for the classifier. Therefore, most of the learning algorithms consider balanced training data, but the minority samples are misclassified frequently. This is because the use of global assessment methods to evaluate the learning algorithm, such as accuracy rate, can be benefit for the majority class. SMOTE [32] is one of the best methods to solve the imbalanced data problem.

SMOTE is used to generate and augment data based on the similarities between the existing minority samples. For each data sample x_k in the minority class $x_k \in C_{\min}$, the k-nearest neighbors/samples are selected. A synthetic sample can be created as follow:

$$x_{\text{new}} = x_k + f_{kl} \cdot \gamma = x_k + \left(\hat{x_{kl}} - x_k\right) \cdot \gamma \qquad (1)$$

where x_k is one of the minority class samples, $\hat{x_{kl}}$ is one of the nearest neighbors for x_k, $\gamma \in [0, 1]$ is a random value, x_{new} is the point/sample along the line joining x_k and $\hat{x_{kl}}$. This implies that a synthetic sample is generated randomly by choosing one of the k-nearest neighbors, $\hat{x_{kl}}$, then multiplying a random number γ with the corresponding feature vector difference f_{kl}, and then adding this vector to x_k [33].

Fig. 1 demonstrates a typical imbalanced data distribution where the red and blue circles represent samples of the majority and minority classes, respectively. A green circle shape highlights the samples produced along with the line segment between x_k and $\hat{x_{kl}}$. These synthetic samples boost the amount of minority samples and thus significantly enhancing the effectiveness of predictions.

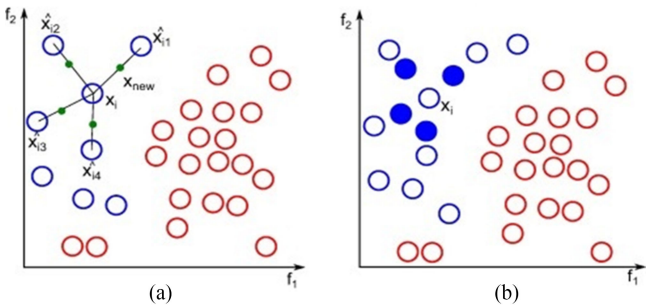

Fig. 1. Example of the SMOT method for binary class. (a) Before applying SMOT. (b) After applying SMOT.

The computational complexity of the proposed SMOTE technique is based on the distances between all instances of the minority class that calculated (*k*). Then, the instances of the majority class that have the smallest distances to those in the minority class are selected (*r*). The overall computational complexity is O(*kr*).

B. New CCSO

The CSO algorithm is a recently developed metaheuristic framework that is based on the intelligent behavior of crows to store their food in a secret place for the purpose of retrieval whenever they need. Crows are regarded as the smartest birds. There is plenty of evidence of crows' cleverness. Crows can remember faces and warn each other when approaching an unfriendly one. They can also use tools, communicate in advanced ways, and remember the hiding place of their food up to several months later.

The four main concepts of the CSO are defined as follows.

1) Crows live in the flocks.
2) Crows keep their intelligent memory of their hiding places of foods.
3) Each crow follows others while doing thievery.
4) Crows are very cautious about thievery and protect their caches by a probability.

Moreover, from the view of optimization, the crows are searching agents, the environment is the search space, each crow position in the environment corresponds to the optimal solution, the quality of food source is an objective (fitness) function, and the position of the best food source of the environment is the global solution of the problem. Based on these intelligent behaviors, a population-based metaheuristic algorithm, CSO, is developed to find the solution of optimization problems [18].

1) CSO Mathematical Model: Suppose that there are *M* crows and *j* denotes the position of the crow at an iteration *t* given by $p^{j,t}$ with $t_{\max}$ being the maximum number of iterations, where $j = 1, 2, \ldots\ldots M$. Each crow has an intelligent memory and remembers the hiding place of food for several months.

Suppose that $m^{j,t}$ is the position of hidden place for crow *j*. *D* is the total number of decision variables. Each crow has an intelligent memory in which the position of its hiding place is memorized.

At iteration *t*, the position of hiding place of crow *j* is shown by $m^{j,t}$. This is the best position that crow *j* has obtained so

far. In the memory matrix of each crow, the position of its best experience has been memorized in each iteration. Crows move in the environment and search for better food sources (hiding places).

Now suppose that the crow *k* at an iteration wish to go to the hiding place $m^{k,t}$ of the food. The crow *j* has decided to chase crow *k* to find the food storing location of the *k*th crow. The following two possibilities cases have the chance to occur.

Case 1: The *k*th crow does not know the chasing of *j*th crow that can lead the *j*th crow to the food position. The new crow position$j$ is determined by the following equation:

$$p^{j,t+1} = p^{j,t} + r_j \cdot \mathrm{fl}^{j,t} \cdot \left(m^{k,t} - p^{j,t}\right) \tag{2}$$

where r_j is the random number generator between $[0, 1]$, and $\mathrm{fl}^{j,t}$ is the FL of crow j at iteration t. Observed that small FLs cause local optima issue, whereas larger flight values provide global optima.

Case 2: In this case, the kth crow is conscious of the chasing of crow j and fooling crow j in this situation by heading to another location in the search space to safeguard the food, which has a random location in the search space. Case 2 can be expressed as follows:

$$p^{j,t+1} = \begin{cases} p^{j,t} + r_j \cdot \mathrm{fl}^{j,t} \cdot \left(m^{k,t} - p^{j,t}\right) & \text{if } \ r_k \geq \mathrm{AP}^{k,t} \\ \text{random position} & \text{otherwise} \end{cases} \tag{3}$$

where $\mathrm{AP}^{k,t}$ indicates the probability of awareness of the crow *k* at iteration *t*. The aim of AP is to control intensification and diversification that occurs due to lower and higher probability awareness values, respectively.

CSO starts with setting the constraints, *D*, $t_{\max}$, *n*, AP, and FL. The memory and position *p* of each crow's are randomly initialized at search space. At the beginning, the crow *p* does not have an experience to hide their food. Thus, they hide their food at initial positions *m*. During the period the algorithm runs, each crow is evaluated using a predefined fitness/cost function. Then, the crows update their positions according to the fitness value, as expressed in (2), and the feasibility of each new position is verified. The crows are updating their memories as follows:

$$m^{j,t+1} = \begin{cases} p^{j,t+1} & \mathrm{Fn}\left(p^{j,t+1}\right) > \mathrm{Fn}\left(m^{j,t}\right) \\ m^{j,t} & \text{otherwise} \end{cases} \tag{4}$$

where $\mathrm{Fn}(.)$ is defined as the objective function. The best position is registered as the optimal solution as soon as the termination criterion is satisfied.

The CSO provides a good equilibrium between exploitation and exploration. This is because CSO is mainly controlled by the only two parameters, FL and AP. By decreasing FL and AP, CSO tends to conduct the search on a local region where a current good solution is found in this region. By increasing the FL and AP, the CSO tends to explore new regions and tends to global search in the search space.

On the other hand, one of the drawbacks of optimization algorithms is setting the parameters as it is a time-consuming task. Therefore, in this work, CSO parameters are tuned in the iteration process using sequence vector of chaotic map.

2) New Binary CSO: In this work, the binary version of CSO is performed using sigmoid transfer function for feature selection. In binary CCSO, the solution pools in a binary form, where the solutions are limited to the binary values $\{0, 1\}$. The agents are transferred from continues to binary space using the following equation:

$$p^{j,t+1} = \begin{cases} 1 & \text{if } \left(s\left(p^{j,t+1}\right) \geq r_{\mathrm{rand}}\right) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $s(p^{j,t+1})$ is sigmoid transfer function that limits the solutions between $\{0, 1\}$ and can be expressed as follows:

$$s\left(p^{j,t+1}\right) = \frac{1}{1 + e^{10(p^{j,t+1} - 0.5)}}. \tag{6}$$

3) New CSO Based on Chaotic Map: Chaotic map is an evolution function used to depict chaotic behavior that can be a discrete or continuous parameter of time. It is provided by chaos theory to study the random and unpredictable deterministic behavior of the system. Usually chaos theory offers the solution to irregular and unpredictable behavior of many nonlinear and complex systems. The chaotic theory with their topologically blend, ergodic, and intrinsic stochastic nature can be applied to balance between diversification and intensification rate [34]. The random variables in CSO are updated by the sequence of circle chaotic vector that influences the optimal solution and convergence rate by updating the crow positions in the search space. Such a combination of chaos with CSO is defined as CCSO. The circle map supplies a straightforward model of the phase-locked iteration. It has two parameters *a* and *b,* parameter *a* is interpreted as the strength of nonlinearity, whereas parameter *b* is interpreted as externally applied frequency. The circle chaotic map sequence is a one-dimensional map, which can be expressed as follows:

$$x_{i+1} = \mathrm{mod}\left(x_i + b - \left(\frac{a}{2\pi}\right)\sin\left(2\pi x_k\right),\ 1\right). \tag{7}$$

The CSO approach is combined with circle chaotic sequences and can be expressed in the following equation:

$$p^{j,t+1} = \begin{cases} p^{j,t} + C_j \cdot \mathrm{fl}^{j,t} \cdot \left(m^{k,t} - p^{j,t}\right) & \text{if } \ C_z \geq \mathrm{AP}^{k,t} \\ \text{random position} & \text{otherwise} \end{cases} \tag{8}$$

where C_j is the obtained value of chaotic map for *j* crow, and C_z is the obtained value of chaotic map for *z* crow.

Algorithm 1 shows the steps of the proposed chaotic CSO for the optimization problem.

4) Objective/Cost Function: At each iteration, each crow position is assessed using a defined fitness function F_n. The data are divided randomly into two different parts, training set and testing set, using the *k*-fold approach. In this study, *k* is set to 3 to guarantee the efficiency and stability of the obtained results. The following equation is adopted in this study as fitness function to evaluate each crow in the search space

$$F_n = \mathrm{Acc} + \alpha \cdot \left(1 - \frac{L_f}{L_t - 1}\right) \tag{9}$$

Algorithm 1: Chaotic Crow Search Optimization Algorithm.

Set initial values of *FL*, *AP*, *M*, and $T_{\max}$
Initialize the crow position *p* randomly
Initialize memory for each crow *m*
Initialize the circle chaotic map *C*.
Initialize counter t = 1
Assign fitness value to each crow.
While ($t <= T_{\max}$)
  For $i = 1{:}M$ // All *M* crows of the flock
    Chaotically choose one of the crows to follow (z)
    If ($C_z \geq AP^{k,t}$)
      $p^{j,t+1} = p^{j,t} + C_j \cdot \ fl^{j,t} \cdot \ (m^{k,t} - p^{j,t})$
    Else
      $p^{j,t+1}$ = random position in search space
    End If
    Apply sigmoid transfer Fn. $s(p^{j,t+1})$. Eq. (5)
    Binarize the solution. Eq. (6)
    t = t + 1
  End For
  Check the feasibility of $p^{j,t+1}$
  Evaluate the new position of crow $F_n(p^{j,t+1})$. Eq. (9)
  Update the crow's memory $m^{j,t+1}$
End while //Terminate when criteria satisfied
Produce the best solution *m* to the problem

where Acc is calculated based on KNN classification accuracy using $K = 5$ and Euclidian distance. KNN is used to determine the goodness of the selected features from the whole data set. L_f is the length of subset selected feature, L_t is the total number of features, α is the weighted factor, and in this study $\alpha = 0.8$. The main criteria is achieved by maximizing the objective function that maximizes the classification accuracy.

5) Computational Complexity of the Proposed CCSO Algorithm: Computational complexity of any algorithm is a key metric for evaluating its run time, which can be defined based on the structure and implementation of the algorithm. The computational complexity of the proposed CCSO approach is mainly dependent on the number of variables (*d*), the number of population (*n*), and the maximum number of iterations (*t*). Since the CCSO utilizes the quicksort mechanism, the computational complexity is O(*n*log*n*) for the best case and O(n^2) for the worst case. The overall computational complexity is O($t$log(n)($n^2 + nd$)).

C. Recognition Based on SOFLP

SOFLP is a recently developed powerful multilayer FRB classifier. SOFLP is a prototype-based approach that defines the suitable local area of influence for each prototype in order to increase the descriptive capacity of the fuzzy rules and avoid overlap. SOFLP is a nonparametric FRB approach that can immediately derive parameters from the data. Furthermore, the prototypes are recognized from the observed data by an offline training method and are used to construct fuzzy rules. This classifier has a good generalization mode and is able to learn

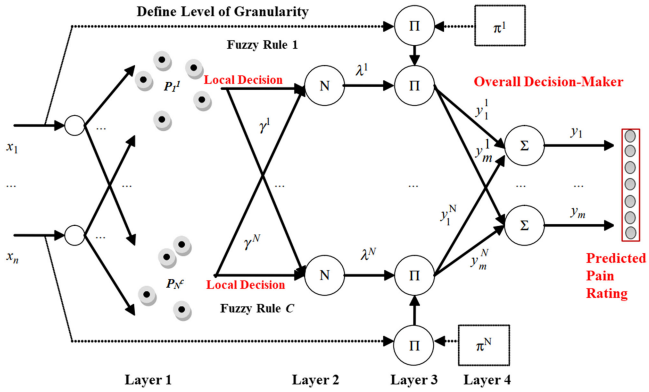

Fig. 2. SOFLP structure of the proposed AnYa FRB.

from the data continually and pursue the evolving data pattern by recursively updating the system structure and metaparameters.

The concept of the SOFLP approach touches the very foundations of the complex system identification, and thus its application domain ranges from simple clustering-based techniques for pattern recognition, image segmentation, etc., to more general modeling, classification, and time-series prediction problems in various applications [35].

The primary issue in FRB is to define membership functions per scalar variable and a very high amount of approximation is required to parameterize all the required real data. Because the real data distributions are often nonlinear, not smooth nor easy to describe. To address this important bottleneck of the FRB systems design, the interpretation of the SOFL approach is introduced in this work [36].

1) SOFLP Layers: The SOFLP approach is represented by four-layer feed-forward neural network, as shown in Fig. 2, which is quite different from the neuro-fuzzy systems, such as adaptive neuro-fuzzy inference system (ANFIS) [37]. In the first layer, no scalar parameterized membership functions are defined; instead, a specified input data sample is compared with all prior data samples per cluster recursively in a computation-effective manner, and the local density of each cluster is conducted.

In the second layer, the density of the respective cluster γ^i is taken as the inputs, and the standardized firing level of the fuzzy rule (which is a membership of the *i*th cluster γ^i) is given as the output. The first and second layers depict the antecedent part of the FRB on a zero-order AnYa FRB introduced recently by Angelov and Yager [22], which is simpler but more complex in the structure than Mamdani and TS types of FRB systems [26].

The antecedent (if) portion of AnYa-type fuzzy rules is simplified to a more compact, objective. and nonparametric vector form compared with the two predecessors without the need to define membership functions [31]. The form of AnYa-type FRB system is as follows:

$$\text{if}\,(x \sim P_1)\,\text{OR}\ (x \sim P_2)\,\text{OR}\ldots\ldots\ldots\text{OR}\ (x \sim P_N)$$
$$\text{Then}\ (\text{class}) \tag{10}$$

where x is the input vector; "$\sim$" denotes similarity (represents the fuzzy degree of satisfaction/membership), $P_i (i = 1\,2\,, \ldots, N)$

is the *i*th prototype of the class, and *N* is the number of prototypes identified from the data samples of this class. The strategy of winner-takes-all is performed to decide the decision of the class labels. For more details of the AnYa-type FRB system, readers are referred to [22] and [31].

The third layer collates the antecedent and the consequent part that represents the local subsystems. Finally, the last layer forms the FRB streamlined system's total output is performed using a weighed summation of local subsystems.

2) SOFLP Phases: The SOFLP approach has three main phases for real-time prediction (offline training, online/ evolving training, and validation).

a) Offline training: The SOFLP approach identifies prototypes individually from each class and forms an AnYa-type FRB on the prototypes recognized per class. These prototypes are recognized based on the information samples' densities and mutual distributions. First, multimodal densities are conducted on all unique data samples u_i, which can be calculated as follows:

$$D_k^{MM}(u_i) = f_i \frac{\sum_{l=1}^{K} \pi_K(x_l)}{2K\pi_K(u_i)}, \qquad i = 1, 2, \ldots.., U_K \tag{11}$$

where K is the number of samples, f_i is the frequencies of occurrence of the unique data sample set u_i, and π_k is cumulative proximity that is acquired from the observed data empirically without prior knowledge or assumptions and can be calculated as follows:

$$\pi_k(x_i) = \sum_{j=1}^{K} d^2(x_i,\ x_j), \qquad i = 1, 2, \ldots\ldots., K \tag{12}$$

where $d^2(x_i,\ x_j)$ refers to the distance between x_i and x_j, which can be any type of distance measure. In this work, we concentrate on performing cosine similarity distance. Because the cosine similarity is free from the "curse of dimensionality," and therefore, it is more effective.

Then, the data samples are ranked in a list denoted by $\{r\}$ in terms of their distances and multimodal density values by finding the largest data sample of multimodal density

$$r_1 = \arg \max_{i=1,2,\ldots\ldots,U_k^c} \left(D_K^{MM}(u_i^c)\right). \tag{13}$$

Afterward, the second element, r_2 is identified as the data sample with the minimum distance to $r_1$as follows:

$$r_2 = \arg \min_{i=1,2,\ldots\ldots,U_k^c-1} d(r_1, u_i^c). \tag{14}$$

The third $\{r\}$ element recognized by r_3 based on the minimum distance to r_2. By repeating the process and until all data samples are chosen, the complete list $\{r\}$ is created, and the multimodal densities of unique samples $\{u\}$ are ranked and denoted by $\{D_k^{MM}(r)\}$.

Afterward, the prototypes p is identified as a local maxima of the ranked multimodal densities $\{D_k^{MM}(r)\}$ and can be expressed as follow:

$$\text{if } \left(D_k^{MM}(r_i) > D_k^{MM}(r_{i+1})\right) \text{ and } \left(D_k^{MM}(r_i) > D_k^{MM}(r_{i-1})\right) \text{ then } (r_i \in \{P\}). \tag{15}$$

Once all the prototypes are identified, filtering operation is performed to filter the less representative prototypes within $\{p\}$. After all the data clusters are created around the current prototypes, one can get the centers of the data clusters, and the multimodal densities at the centers are calculated.

The level of granularity *L*th ($L = 1, 2, 3, \ldots M$) is obtained in an offline manner from the *c*th class, which is a local influential area around each data sample. Finally, the most representative prototypes of the *c*th class denoted by $\{p_c\}$ are selected out from the centers of the existing data clusters.

b) Online/Evolving training: The SOFLP classifier continues to update its parameters and structure with a new data during the online training phase. Similar to the offline training phase, the fuzzy rules of different classes are updated individually. Also, recursive calculation of the following expressions with cosine distance are performed:

$$\mu_k = \frac{K-1}{K}\mu_{K-1} + \frac{1}{K}x_K, \quad \text{where } \mu_1 = x_1 \tag{16}$$

$$X_K = \frac{K-1}{K}X_{K-1} + \frac{1}{K}x_K^T x_K, \quad \text{where} \quad X_1 = x_1^T x_1 \tag{17}$$

$$\sum_K = \frac{K}{K-1}\left(X_K - \mu_K^T \mu_K\right) \tag{18}$$

where the metaparameters $\sum_K$is a covariance matrix, x_K are the data set samples, and μ_K is the global mean that can be updated recursively based on the new data. After updating the classifier's metaparameters, the AnYa-type fuzzy rules will be updated accordingly, and the fuzzy classifier will be ready to process and classify the next data sample.

c) Validation: In this phase, as shown in Fig. 2, the fuzzy classifier offers the decision-making. During the validation phase, for a particular testing data sample, denoted by x, each AnYa-type fuzzy rule will have a firing strength denoted by the local decision-make $\lambda^c(x)$, which is determined as follows:

$$\lambda^c(x) = \max_{p\epsilon\{p\}} \left(e^{-d^2(x,p)}\right), \quad c = 1, 2, \ldots\ldots.., C. \tag{19}$$

Accordingly, based on the *C* firing strengths of the *C* fuzzy-rules (one per rule), the overall decision-maker decides the x label using the winner-takes-all principle as follows:

$$\text{PCL} = \arg \max_{c=1,2,\ldots.C} \left(\lambda^c(x)\right) \tag{20}$$

where PCL is the prediction class label.

3) Tuning Level of Granularity: Solving a complex problem involves dividing the problem into information granules and managing each granule as a whole through a granular computing strategy. The granularity idea was introduced by Yao *et al.* [38] to a multilevel view of information granulation, which permeates human reasoning and has a significant effect in any field involving both human and machine-oriented problem solving.

Granular computing has largely contributed to the fuzzy theory as a paradigm of problem solving and information processing inspired by humans [39]. In general, the higher level of granularity is chosen, the more prototypes (more details)

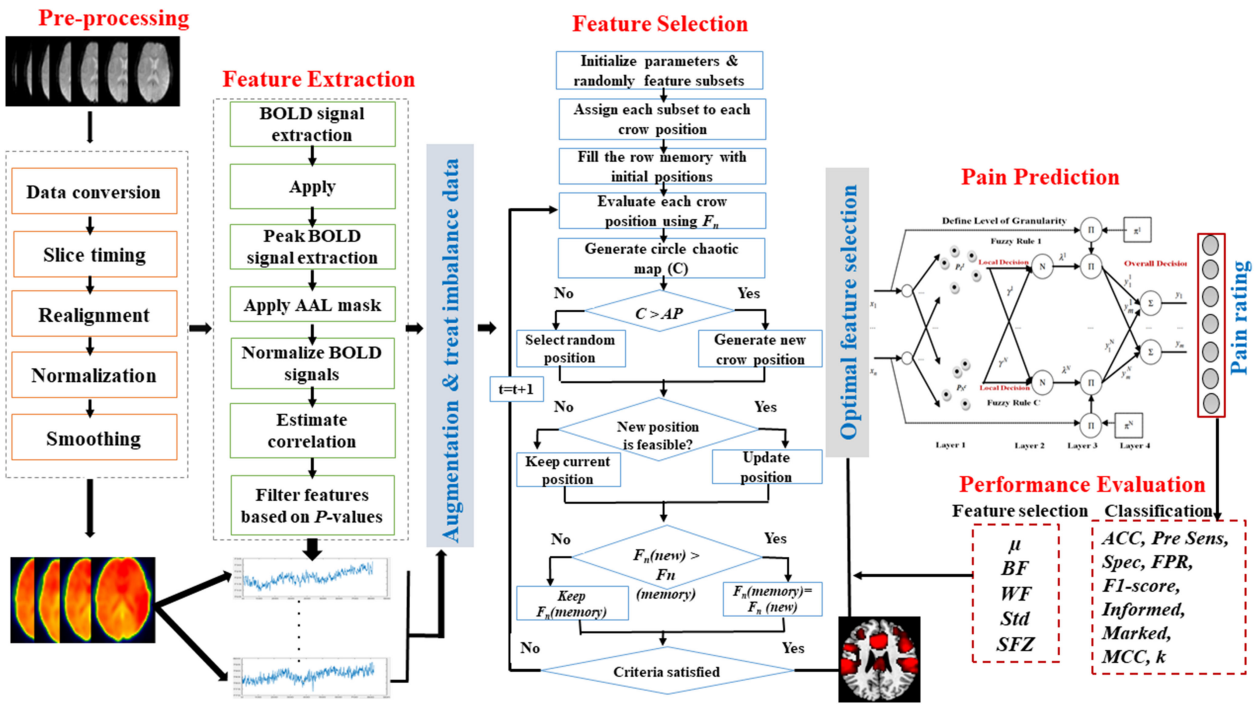

Fig. 3. Pipeline and phases of the proposed CCSO–SOFLP approach for multiclassification pain prediction.

are extracted from the data, and the SOFLP approach achieves a higher performance. At the same time, more computational and memory resources may be consumed by the SOFLP, and overfitting may be occurred. On the other hand, with low level of granularity, the SOFLP only learns the coarse information from training. Although it will decrease the time consumption, its performance may be influenced due to the loss of fine information from the data.

The derived hierarchical collection of granules can be used to build a fuzzy model committee, which offers a good equilibrium between interpretable representation and precise approximation.

4) Computational Complexity of the SOFLP Approach: The computational complexity of this approach is mainly dependent on two parts: premise trainable variable part and consequent part per rule. In the premise trainable variable part, it depends on the number of input variables (d) which are observed from the CCSO approach— each input d is partitioned into (p) prototypes—and the recursive trained variables with local density (m). According to these variables, the total number of premise trainable part is (dpm). Similarly, the consequent part per rule is $(p + 1)$ and number of rules in the system is (c^d). According to this, total number of consequent part is $((p + 1)c^d)$. The overall computational complexity is $O(dpm((d + 1)c^d))$.

III. Proposed CCSO–SOFLP for fMRI-Based Pain Prediction

A. Experiment and Data

fMRI data were obtained from 31 healthy participants (20 females, age, 22.1 ± 2.0, ranging from 19 to 24 years). Participants reported no history of chronic pain, psychiatric, or neurological disorders. The local ethics committee endorsed the experimental procedures (Approval no. SWU20140607). All participants gave written informed permission. They were acquainted with the experiment paradigm prior to the experiment. The fMRI data were obtained using a Siemens 3.0 Tesla Trio scanner with a standard head coil. The following parameters were used to acquire functional images with echo planar imaging sequence: 255-mm-thick slices and 0.5-mm interslice gaps, repetition time (TR) = 1500 ms, echo time (TE) = 29 ms, filed of view = 192×192 mm^2, 64×64 matrix, $3 \times 3 \times 3$ mm^3 voxels, and flip angle = 90. A high-resolution T1-weighted structural image (1 mm^3 isotropic voxel MPRAGE) was acquired after functional imaging.

We produced 10 laser pulses in each of the 4 stimulus energies (E1: 2.5 J, E2: 3 J, E3: 3.5 J, E4: 4 J) during the fMRI data collection, making a total of 40 trials. The order of stimuli at different energies was pseudorandomized. The interstimulus interval (rectangular distribution) ranged from 27 to 33 s randomly. The subject rated the intensity of pain perception elicited by the laser stimulus using a visual analogue scale ranging from 0 ("no pain") to 10 ("the worst pain").

B. Proposed CCSO–SOFLP Pipeline

In this section we present a new hybrid optimization CCSO–SOFLP approach to pain intensity decoding. Fig. 3 shows the pipeline view of the proposed CCSO–SOFLP approach for the pain intensity prediction. Basically, the whole CCSO–SOFLP approach consists of the following main phases.

1) fMRI Data Preprocessing: The aim of preprocessing is to remove noise and artifacts from the fMRI data. In this study, the fMRI data were preprocessed using Statistical Parametric

Mapping (SPM12) [40] and followed the following procedures: slice timing and head motion correction. The data were realigned, normalized to the Montreal Neurological Institute (MNI) space (voxel size $= 3 \times 3 \times 3$) by mapping T1-weighted structural images to the MNI template, and smoothed to reduce the effects of bad normalization and to increase the signal-to-noise-ratio using 8-mm full-width half-maximum criterion and Gaussian kernel. The default mask was used with a threshold of 50%. Finally, the Friston 24-parameter model was used to regress head movement impacts and to regress out several nuisance signals from the data. A high-pass filter (cut-off frequency $=$ 1/128 Hz) was introduced into the blood-oxygen-level-dependent (BOLD) time sequence to remove low-frequency drifts. BOLD responses were modeled as a series of events using a stick function and ratings were included as a parametric modulator of each stimulus, which were then convolved with a canonical hemodynamic response function.

2) Feature Extraction: BOLD responses have been normalized by subtracting their baseline values and dividing the results at stimulus onset [41]. The following equation was modeled to normalize BOLD signals between the maximum BOLD signals $(\mathrm{Max_{BOLD}})$ and the BOLD signals at stimulus onset $(\mathrm{Stm_{Onset}})$

$$\mathrm{Norm_{BOLD}} = (\mathrm{Max_{BOLD}}) - \mathrm{Stm_{Onset}} \,/\, \mathrm{Stm_{Onset}}. \quad (21)$$

The maximum BOLD $\mathrm{Max_{BOLD^{4th}}}$ responses at the fourth scan after stimulus onset were extracted as fMRI features for prediction of subjective pain ratings as expressed in the following equation:

$$\mathrm{Max_{BOLD^{4th}}} = |(\mathrm{Stm_{Onset}} \,/\mathrm{TR})| + 4 \quad (22)$$

where TR is the repetition time that represents cycle time between corresponding points, TR in this study was 1.5 s.

The dimension matrix of features extracted was $1240 \times 56\,260$ voxels. Each subject had 40 trials, and 31 subjects had 1240 trials in total with number of voxels is 56 260. To normalize the maximum BOLD signals at fourth scan after stimulus onset $(\mathrm{Norm_{BOLD}})$. The following equation was modeled to normalize BOLD signals between the maximum BOLD signals $(\mathrm{Max_{BOLD^{4th}}})$ and the BOLD signals at stimulus onset $(\mathrm{Stm_{Onset}})$

$$\mathrm{Norm_{BOLD}} = (\mathrm{Max_{BOLD^{4th}}}) - \mathrm{Stm_{Onset}} \,/\, \mathrm{Stm_{Onset}}. \quad (23)$$

After fMRI was normalized, the fMRI voxels were filtered based on the correlation coefficient to rank the features and to select the voxels with the top 4% highest *P*-value.

3) Augmentation of Imbalanced fMRI Data: We had 11 classes of pain intensity prediction from 0 to 10 where each class had the following distributions of examples: Class 0: 59 trials, Class 1: 128 trials, Class 2: 119 trials, Class 3: 136 trials, Class 4: 132 trials, Class 5: 106 trials, Class 6: 168 trials, Class 7: 152 trials, Class 8: 122 trials, Class 9: 82 trials, and Class 10: 36 trials. The number of samples in different classes was largely different. In this case, the classification accuracy rate can be degraded and biased to the majority classes because the imbalanced data distribution with the classes have higher majority instances than others. The best practical in this situation is to apply the SMOTE technique to generate synthetic samples from the minority classes, as described in Section II-A.

4) Feature Selection: CCSO was used to determine the significant fMRI voxels in order to obtain the subset of pain-related features from high-dimensional neuroimaging fMRI data. The chaotic vector was initialized in the form of circle chaotic map that was used in the iteration process as a random chaotic motion instead of basic random behavior in the CSO algorithm to balance between the exploration and exploitation rates. CSO then generated the crow positions following chaotic motion as well as assigned fitness value to each crow based on KNN as an objective function. The KNN was used as a cost function for the proposed CCSO algorithm and worked as a part of the optimizer to assess every crow position in the search space. The main criterion was satisfied by maximizing the objective function. The good solutions were stored in CSO intelligent memory after the position and fitness were evaluated and updated in every iteration until the criterion was satisfied to attain the global best fitness function as the global solution for the problem.

5) Multiclassification of Pain Intensity: The SOFLP multiclassification approach was established to describe the relationship between the levels of pain intensity perception and fMRI patterns. SOFLP is able to learn consistently from newly arrived data samples and only store the key information in memory, which can be considered as memory- and computation-efficient. Furthermore, the proposed approach is nonparametric and work without prior assumptions for parameters and memberships. SOFLP depends highly on the level of granularity (information details) to increase the performance and decrease the computational time. SOFLP has mainly three phases (offline training, online training/ evolving, and validation). In the offline training phase, the examples are labeled, and the SOFLP classifier, as shown in Fig. 2, identifies prototypes from each class separately (i.e., each class has more than one prototype). SOFLP automatically identifies the prototypes from the observed fMRI pain data and forms clusters without prior information. Thus, for a training data set with labels, independent if…then AnYa-type FRB subsystems are generated (one per class) in parallel, as shown in (10). Once the training process is finished, each subsystem generates one AnYa-type fuzzy rule corresponding to its own class based on the identified prototypes. This AnYa-type fuzzy rule is interpreted as a single prototype connected by the "OR" operator. As a result, a massive parallelization is possible. During the online training stage, the FRB system was recognized from the offline training process, which was subsequently updated with the new samples of the fMRI data. During the validation stage, the SOFLP classifier provided the decision of pain intensity based on firing strength of AnYa-type fuzzy rules with the "winner-takes-all" strategy.

6) Performance Evaluation: The efficiency of the proposed CCSO–SOFLP approach was assessed using various statistical measures for feature selection and multiclassification tasks. These measures included: mean fitness (μ), best fitness (BF), worst fitness (WF), standard deviation (Std), and selected feature size (SFZ) for feature selection task [14], accuracy (ACC), precision (Pre), sensitivity (Sens), specificity (Spec), false positive rate (FPR), F1-score, informedness (Informed), markedness (Marked), Matthews correlation coefficient (MCC), and *k*-kappa coefficient (k) for classification task [42]. The efficiency of the proposed approach depends on the tenfold

TABLE I
CCSO PARAMETER SETTING

Parameter	Value	Description
M	5	No. of crows
AP	0.1	Awareness probability
Fl	2	Flight length
L	0	Lower bound
U	1	Upper bound
T_{Max}	100	Maximum iteration
D	56260	Dimension of brain voxels

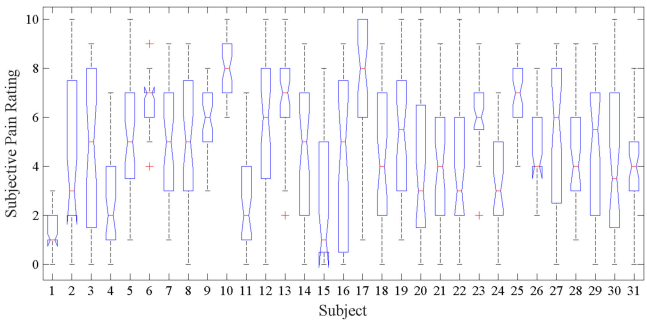

Fig. 4. Subjects pain intensity distribution.

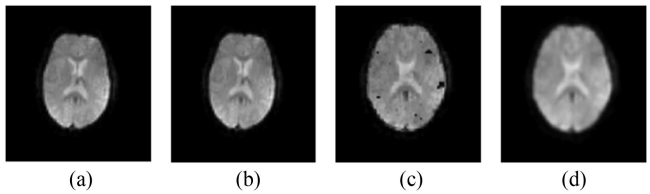

Fig. 5. Results of the preprocessing steps. (a) Original volume. (b) After realignment. (c) After normalization. (d) After smoothing.

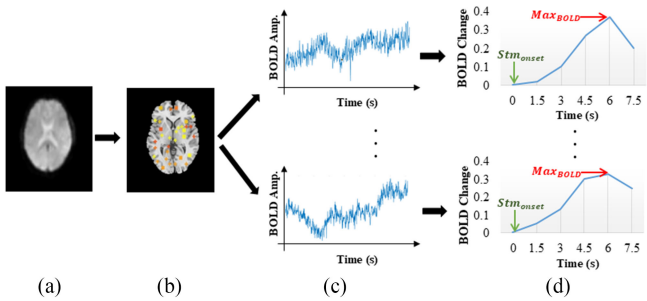

Fig. 6. Results of the BOLD signal extraction and feature analysis steps from one subject. (a) Preprocessed volume. (b) Voxels represent pain perception. (c) BOLD time series extraction. (d) Maximum BOLD signal at the fourth scan after stimuli onset extraction.

CV to reduce the variance and to decrease overfitting and underfitting.

IV. RESULTS

A. Computer System Description and Parameter Setting

The proposed model was developed and tested in MATLAB R2018b on an IntelCoreTM I7-8700 with 3.2 GHz CPU and 64 GB RAM.

CCSO started with setting adjustable parameters and randomly initialized crow positions (solution) in the search space. Each position represented a feature subset with different numbers of features and different lengths. The initial parameter setting of CCSO is presented in Table I.

B. Behavioral Data Analyses

As we mentioned, one big challenge in fMRI-based prediction is the imbalanced data, which means different classes have largely different numbers of sample. Fig. 4 illustrates the imbalanced data by showing the rated pain intensity scales of 31 subjects (each having 40 rated scales). As we can see, in response to the same set of painful stimuli, the subjects had different rates of pain perception. Average pain rates varied between 2.63 ± 2.11 and 7.63 ± 2.23. Using Pearson's correlation coefficient (PCC) analyses, we found no significant relation between age and weight with averaged pain ratings (age and rating: $R = -0.175$, $P = 0.356$; weight and rating: $R = 0.0978$, $P = 0.607$, respectively), where R is the PCC between the fMRI features and the rating class labels, and P represents the p-value of PCC.

In addition, there was no significant distinction between males and females in their average pain scores (two-sample t-test, $P = 0.428$). Therefore, weight and gender cannot explain the relationships between pain perception and stimuli. Moreover, the number of subjects in this study is still small, and a large-sample study should be carried out to generate outcomes that are more potent.

C. fMRI Preprocessing and Feature Extraction

The outcomes of the fMRI preprocessing steps are shown in Fig. 5: realignment, normalization, and smoothing. Fig. 6 shows the time-series BOLD signal extraction and local feature analysis based on maximum BOLD signal at the fourth scan (1.5 s TR) after stimuli onset, which was located around 6 s, as shown in Fig. 6(d), whereas the deactivation regions were located around fifth scan (7.5 s).

D. Augmentation of Imbalanced fMRI Data

The number of subjects and the number of trials per individual in this experiment were not large, which limited the statistical power of the results. In this study, we analyzed fMRI data for 31 subjects. Increasing the number of subjects will make the data contain more possible variations across-subjects and within-subjects, which will be carried out to produce more significantly powerful results. Furthermore, more painful stimuli is needed to be delivered in a wider range of intensities so that the pain rating evoked can also have a wider variety, which can make class labels of the classification model more complete and lead to a more accurate and general classification model. We had 11 class labels for pain intensity from 0 to 10. The distribution of these classes had problem of imbalance, which means it had largely different numbers of samples (fMRI images) for different classes (pain scales) and degraded the accuracy of the prediction model. Therefore, this article incorporates the SMOTE technique based on KNN to overcome the limitation of samples and perform data

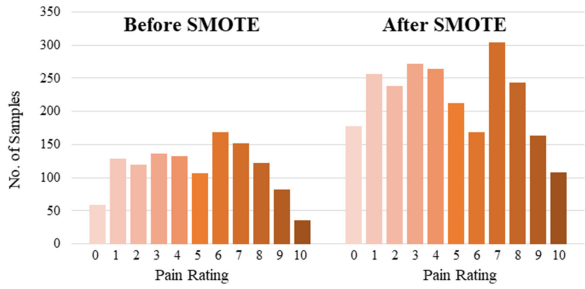

Fig. 7. Results represent the number of instances for each class label before and after the SMOTE technique.

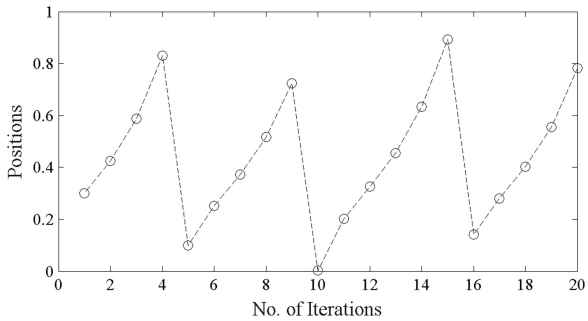

Fig. 8. Behavior of CSO based on circle chaotic map for 20 iterations.

TABLE II
RESULTS OF THE PROPOSED CCSO-BASED SOFLP APPROACH

Appro.	μ	Std	BF	WF	AAS	Time/s	ACC
CSO	0.7021	0.0753	0.7488	0.5224	0.2492	2.16E+02	0.8687
CCSO	0.7319	0.0622	0.7555	0.5224	0.2249	1.74E+02	0.8917

augmentation. In order to increase the number of samples in the pain intensity data set, synthetic data samples were generated from the minority classes, as shown in Fig. 7.

E. CCSO-Based Feature Selection

To identify the brain regions that significantly contribute to pain perception, the CSO algorithm and circle chaotic map were used. Fig. 8 shows the behavior of circle chaotic map for 20 iterations of the proposed CCSO algorithm. The behavior of CCSO is reasonably simple, and the crows (agents) are moved chaotically instead of stochastic behavior in the search space, as shown in Fig. 8.

Table II shows the comparison results of the proposed CCSO and CSO in terms of μ, Std, BF, WF, SFZ, and ACC. As we can see, the proposed CCSO approach attained the better accuracies in terms of ACC ($\approx$0.89), AAS ($\approx$0.22), BF ($\approx$0.75), and μ ($\approx$0.73), all of which were better than the standard CSO for pain intensity prediction. In addition, the proposed CCSO time consumption (CPU average) was much lower than the CSO algorithm. This proves that the proposed CCSO is robust and can achieve the optimal solution in fewer iterations with less time consumption.

The convergence speed evidence is provided in Fig. 9. The CCSO showed a strong robustness and had faster convergence

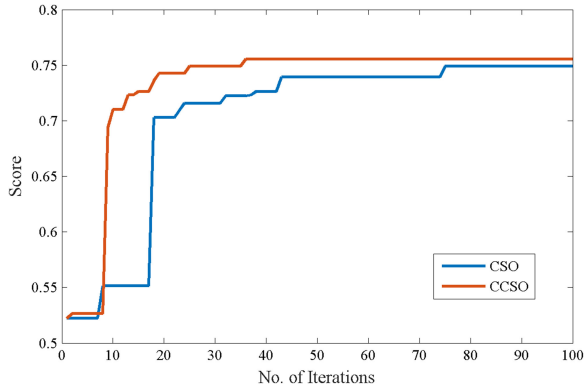

Fig. 9. Convergence curve using CSO and CCSO.

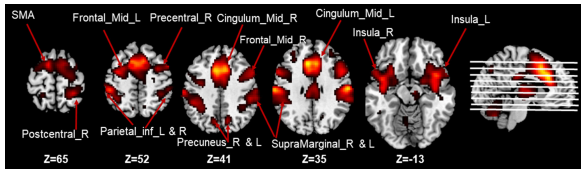

Fig. 10. Brain regions with predictive pain-evoked BOLD responses extracted by the proposed approach at the fourth scan after stimulus onset.

speed to find the optimal or near optimal solutions in less than 40 iterations for the fMRI subjects.

The identified pain-related brain patterns for a single-subject data using laser-evoked BOLD responses are shown in Fig. 10 based on the CCSO–SOFLP approach. This figure shows the brain decoding and the region of interests (ROIs) that were most affected by stimuli. The general linear model was used to assess the laser-evoked BOLD activation voxels by CCSO–SOFLP and to extend the cluster neighbor voxels [10], [43], [44]. The CCSO analyses revealed that the BOLD responses within a wide range of brain regions are predictive of pain perception, as shown in Fig. 10. These predictive regions are the primary (S1) and secondary (S2) somatosensory cortex, thalamus, anterior cingulate cortex, supplementary motor area, mid-cingulate cortex, cerebellum, and insula, frontal, post-central, supra-marginal, etc. Among these predictive regions, SOFLP multiclassifier approach is proposed for pain rating level. From the results, we can demonstrate that the CCSO is able to solve the problems of high dimensionality and multicollinearity, which are typical in neuroimaging data.

F. SOFLP-Based Pain Classification

In this section, the subset of the most predictive features of pain perception were selected. These features were feed into the SOFLP approach for pain-level prediction. The proposed pain classification approach was trained to decode single-trial pain perception intensity from pain-related patterns recognized by the fMRI. We performed a multiclass pain intensity classification based on the SOFLP approach to predict the level of pain perception within the range between 0 and 10 for each trial (40 trials/subject). Ten-fold CV was used to validate the strategy of the pain decoding by between-subject and within-subject levels.

TABLE III
RESULTS OF THE PROPOSED CCSO–SOFLP APPROACH USING TENFOLD CV AND DIFFERENT GRANULARITY (L) LEVELS

L	*CV1*	*CV2*	*CV3*	*CV4*	*CV5*	*CV6*	*CV7*	*CV8*	*CV9*	*CV10*	**Average**
1	0.2739	0.2158	0.2324	0.2917	0.2739	0.2583	0.2241	0.2531	0.2531	0.2375	0.2514
2	0.3444	0.3112	0.3278	0.3195	0.4025	0.4025	0.4000	0.3125	0.3542	0.3734	0.3548
3	0.4772	0.5394	0.5375	0.4896	0.4606	0.4730	0.5167	0.5145	0.444	0.5417	0.4994
4	0.7303	0.6763	0.6432	0.6792	0.6473	0.6971	0.6224	0.6667	0.5851	0.6375	0.6585
5	0.8083	0.8465	0.8299	0.8417	0.8174	0.8174	0.8333	0.8008	0.7925	0.7842	0.8172
6	0.8506	0.8625	0.8542	0.8672	0.8672	0.8423	0.8340	0.8382	0.825	0.8174	0.8459
7	0.8465	0.9000	0.9170	0.8714	0.8880	0.8423	0.8631	0.8755	0.8917	0.8583	0.8754
8	0.8833	0.8838	0.8631	0.8792	0.8921	0.8340	0.8797	0.8921	0.9004	0.9042	0.8812
9	0.8963	0.8708	0.8797	0.8797	0.8838	0.9170	0.8833	0.9004	0.8382	0.8542	0.8803
10	0.8714	0.8755	0.8838	0.8833	0.8875	0.8465	0.8797	0.8838	0.8792	0.8797	0.877
11	0.8875	0.8838	0.8921	0.8921	0.8714	0.8625	0.8631	0.8963	0.8625	0.9087	0.8820
12	0.8917	0.9000	0.8672	0.8797	0.8583	0.8963	0.8963	0.8797	0.9129	0.8506	0.8833
13	0.8921	0.888	0.8506	0.85	0.8838	0.8589	0.9212	0.888	0.875	0.8917	0.8799

TABLE IV
RESULTS OF THE PROPOSED CCSO–SOFLP APPROACH USING DIFFERENT STATISTICAL MEASURES AND DIFFERENT GRANULARITY (L) LEVELS

L	*ACC*	*Sens*	*Spec*	*Pre*	*FPR*	*F1_score*	*MCC*	*Informed*	*Marked*	*Time/s*
1	0.2514	0.2740	0.9255	0.2793	0.0745	0.2502	0.1947	0.1995	0.1537	24.9531
2	0.3548	0.3765	0.9355	0.3777	0.0645	0.4102	0.3067	0.3120	0.2661	27.2500
3	0.4994	0.5169	0.9497	0.5068	0.0503	0.5118	0.4559	0.4666	0.4368	27.4375
4	0.6585	0.6714	0.9656	0.6457	0.0344	0.6862	0.6203	0.6370	0.6187	31.5000
5	0.8172	0.8096	0.9816	0.7991	0.0184	0.7892	0.7792	0.7912	0.8025	33.8125
6	0.8459	0.8370	0.9845	0.8110	0.0155	0.7992	0.8046	0.8215	0.8385	34.4219
7	0.8754	0.8664	0.9874	0.8607	0.0126	0.8484	0.8440	0.8538	0.8628	34.5938
8	0.8812	0.8702	0.9881	0.8617	0.0119	0.8517	0.8473	0.8583	0.8745	37.7500
9	0.8803	0.8709	0.9879	0.8634	0.0121	0.8543	0.8493	0.8588	0.8717	35.8906
10	0.8770	0.8680	0.9876	0.8653	0.0124	0.8473	0.8450	0.8556	0.8658	36.5625
11	0.8820	0.8708	0.9881	0.8594	0.0119	0.8561	0.8481	0.8589	0.8724	39.5781
12	0.8833	0.8755	0.9882	0.8779	0.0118	0.8595	0.8569	0.8638	0.8710	38.2344
13	0.8799	0.8685	0.9879	0.8626	0.0121	0.8485	0.8454	0.8564	0.8745	37.2500

Solving the complex nonlinear fMRI pain prediction problem via a granular computing can provide a good balance between interpretable representation and precise approximation. The proposed approach relies on the granularity level in order to improve the efficiency and reduce the computational time. Table III shows the results based on different levels of granularity (L) that lead to fine details for the SOFLP approach. From Table III and Fig. 11, we can see that the best accuracies can be achieved using 12th L, which ensured that the SOFLP classifier can learn from sufficient details. Furthermore, we compared the result of the proposed approach at (12th L) with the results of different granule levels in ten independent runs, and then we used paired-sample t-test to compare their accuracy with that of $L = 12$. It was showed that the results became more stable after the granular level

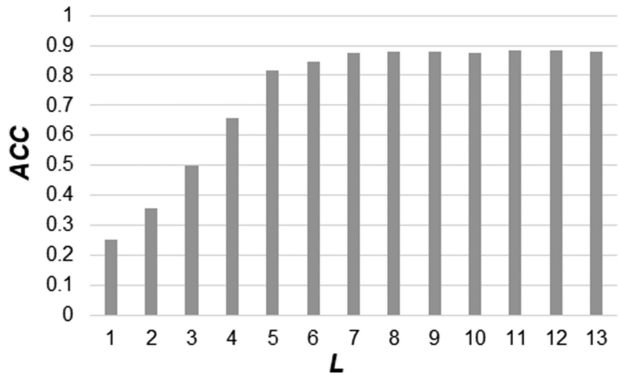

Fig. 11. Comparison results of the proposed approach using different levels of granularity (L).

TABLE V
RESULTS OF PAIN PERCEPTION WITHIN EACH SUBJECT (31 SUBJECTS) USING THE CCSO–SOFLP APPROACH

Subj.	*ACC*	*Sens*	*Spec*	*FPR*	*Informed*	*Marked*
1	1	1	1	0	1	1
2	0.9444	0.9500	0.9933	0.0067	0.9433	0.9026
3	0.8333	0.8667	0.982	0.018	0.8487	0.8052
4	0.6875	0.7619	0.9429	0.0571	0.7048	0.5067
5	0.8889	0.900	0.9871	0.0129	0.8871	0.8405
6	0.8750	0.9167	0.9762	0.0238	0.8929	0.8259
7	0.7647	0.7500	0.9658	0.0342	0.7158	0.7078
8	0.8333	0.800	0.9820	0.0180	0.7820	0.7839
9	0.8824	0.8333	0.9744	0.0256	0.8077	0.8273
10	0.9412	0.9000	0.9833	0.0167	0.8833	0.9030
11	0.8235	0.8333	0.9738	0.0262	0.8071	0.7580
12	0.9444	0.9091	0.9943	0.0057	0.9034	0.9150
13	0.8235	0.8889	0.9755	0.0245	0.8644	0.7123
14	0.9444	0.900	0.9941	0.0059	0.8941	0.9477
15	0.6250	0.8125	0.9457	0.0544	0.7582	0.4923
16	0.8824	0.8889	0.9847	0.0153	0.8735	0.8407
17	0.9412	0.9630	0.9921	0.0079	0.9550	0.8952
18	0.8235	0.8833	0.9800	0.0200	0.8633	0.7372
19	0.6875	0.7500	0.9636	0.0364	0.7136	0.5900
20	0.9412	0.9500	0.9933	0.0067	0.9433	0.9121
21	0.7500	0.7667	0.9703	0.0297	0.7369	0.5909
22	0.8824	0.9242	0.9883	0.0117	0.9125	0.8367
23	0.8750	0.9286	0.9780	0.0220	0.9066	0.8123
24	0.9412	0.9375	0.9922	0.0078	0.9297	0.9445
25	0.9412	0.9444	0.9861	0.0139	0.9306	0.8925
26	0.6250	0.7667	0.9172	0.0828	0.6839	0.5528
27	0.7647	0.7778	0.9713	0.0287	0.7491	0.6693
28	0.6667	0.6667	0.9569	0.0431	0.6236	0.5511
29	0.7857	0.7000	0.9769	0.0231	0.6769	0.7672
30	0.7647	0.8000	0.9750	0.0250	0.7750	0.7159
31	0.8235	0.8750	0.9712	0.0288	0.8462	0.6674
Mean	0.8302	0.8515	0.9756	0.0244	0.8271	0.7635

$L = 7$ and there was only slightly difference between them [p-value ($L = 1-6$) < 0.004 and p-value ($L = 7-12$)> 0.352].

Table IV shows the results of the proposed approach under different statistical measures and different levels of granularity. The classification performances were significantly estimated by ACC, Pre, Sens, FPR, Spec, F1-score, Informed, Marked, and MCC. From the results, the highest accuracy can be achieved (ACC ≈ 0.8833) using 12th L with low time consumption $\approx$38.2344 s (CPU time average/second). Also, the best

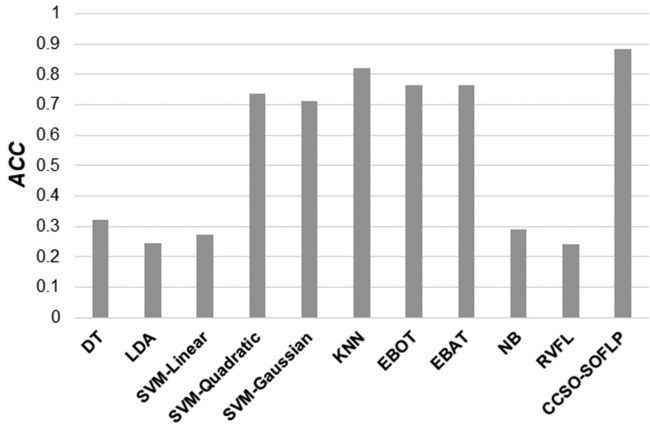

Fig. 12. Graphical representation for comparison between the performance of CCSO–SOFLP and other ML methods.

or satisfactory values for other measures were Sens$\approx$0.8755, Spec$\approx$0.9882, Pre$\approx$0.8779, and F1-score$\approx$0.8595, which represented a perfect precision and recall even if the data were unbalanced. MCC$\approx$0.8569 and Informed$\approx$0.8638 measured how the system was informed about positives and negatives, and Marked$\approx$0.871 measured the trust worthiness of positive and negative predictions by the model. Therefore, SOFLP could be potentially used as a core part of the proposed fMRI-based pain prediction in clinical practice.

Table V shows the accuracy results within-subject performance for 31 subjects to demonstrate significant variations in BOLD responses for each subject and to examine whether the proposed approach can achieve consistently good performance for all subjects. From the results, we can see the proposed CCSO–SOFLP approach had high classification accuracy and achieved good performance for all subjects and showed variation of BOLD response between subjects in pain intensity perception.

Taking *Accuracy* as an example, almost all of the subjects had accuracy larger than 80% and some subjects had accuracy of 100%. It shows that the performance of the proposed method is consistent and robust.

G. Comparison With Different ML Techniques

This section provides the comparison results between the proposed CCSO–SOFLP approach and different ML methods for pain-level perception, including naïve Bayes (NB), support vector machine (SVM) with quadratic and Gaussian kernels, linear discriminate analysis (LDA), KNN, decision tree (DT), Random Vector Functional Link (RVFL) [45], ensemble bagged trees (EBAT), and ensemble-boosted trees (EBOT). Furthermore, one of the main objectives of this article is to make a consistent comparison between the classification accuracy obtained by the proposed model and the accuracy obtained by other ML algorithms. Hence, in this section, comparison with other algorithms are shown after using CCSO feature selection on the same pain intensity data set using ten-fold CV.

The comparison results are shown in Table VI and Fig. 12. It can be seen that the proposed CCSO–SOFLP approach can achieve the highest ACC ($\approx$0.8833), and its Sens ($\approx$0.8755),

TABLE VI
COMPARISON RESULTS USING DIFFERENT MEASURES OF DIFFERENT ML METHODS AND THE PROPOSED CCSO–SOFLP APPROACH BASED ON TENFOLD CV FOR FMRI PAIN INTENSITY PERCEPTION

Methods	*ACC*	*Sens*	*Spec*	*Pre*	*FPR*	*F1_score*	*MCC*	*Informed*	*Marked*
DT	0.3223	0.3206	0.9317	0.314	0.0683	0.3123	0.2473	0.2523	0.2614
LDA	0.246	0.247	0.9239	0.2433	0.0761	0.2444	0.1688	0.1708	0.1703
SVM-Linear	0.2733	0.2572	0.9262	0.2837	0.0738	0.2576	0.1924	0.1834	0.191
SVM-Quadratic	0.736	0.7304	0.9733	0.7214	0.0267	0.7229	0.6983	0.7038	0.7098
SVM-Gaussian	0.712	0.6853	0.97	0.5105	0.03	0.3763	0.4437	0.6554	0.4092
KNN	0.819	0.8007	0.9818	0.8158	0.0182	0.7735	0.7673	0.7825	0.7985
EBOT	0.7653	0.7573	0.9762	0.7593	0.0238	0.7498	0.7313	0.7335	0.7286
EBAT	0.765	0.7573	0.9762	0.7593	0.0238	0.7498	0.7313	0.7335	0.7286
NB	0.2895	0.3058	0.9294	0.371	0.0706	0.2756	0.2432	0.2352	0.1742
RVFL	0.2424	0.2272	0.9231	0.2532	0.0769	0.6402	0.1681	0.1504	0.1606
CCSO-SOFLP	0.8833	0.8755	0.9882	0.8779	0.0118	0.8595	0.8569	0.8638	0.871

Spec (≈0.9882), Pre (≈0.8779), FPR (≈0.0118), and F1-score (0.8569) were also the best or close to the best results. In conclusion, the proposed CCSO–SOFLP approach is superior to conventional ML methods.

V. CONCLUSION

In this article, we proposed a novel dynamic multiclassification pain intensity approach, CCSO-SOFPL. First, the CCSO method can select the optimal feature set, which contains a small subset of the most discriminative features, from high-dimensional, nonlinear, and complex fMRI data for pain intensity decoding. Second, the SOFLP method is used for multiple-class prediction of pain levels. SOFLP is adopted in this article because as a powerful multilayer FRB method, it is fully automatic, self-organizing, highly parallel, memory efficient, and self-evolving. The advantages of SOFLP make it particularly suitable for fMRI-based pain prediction, because we need a fast, efficient, and automatic classifier to handle high-dimensional and complex fMRI data. The evolving strategy is utilized in this approach to make the system dynamically learn from new cases to keep its stability and to avoid retraining from scratch. Although the pain data set is not large and evolving in this study, it could be large and incremental in clinical practice. Suppose more data from patients are available in hospitals, we need not train a new SOFLP classifier but can continuously update metaparameters and structures of the classifier based on newly arrived data. Therefore, SOFLP could be potentially used as a core part of the proposed fMRI-based pain prediction in clinical practice.

The efficiency of the proposed approach was evaluated on a real fMRI data set with pain ratings acquired in a laser-evoked pain experiment. Results showed that the CCSO–SOFLP approach can identify brain patterns that were nonlinearly correlated with pain perception and achieved significantly higher pain classification accuracy than conventional ML techniques.

In summary, the proposed CCSO–SOFLP approach is a powerful feature selection and prediction framework for fMRI-based pain prediction, which can overcome the limitations of conventional feature selection and ML techniques. Therefore, the proposed approach has a great potential to be used to develop neuroimaging-based pain assessment tools for clinical uses.

REFERENCES

[1] J. D. Loeser and R. D. Treede, "The Kyoto protocol of IASP basic pain terminology," *Pain*, vol. 137, no. 3, pp. 473–477, 2008.
[2] H. Breivik *et al.*, "Assessment of pain," *Brit. J. Anaesthesia*, vol. 101, no. 1, pp. 17–24, 2008.
[3] E. Schulz, A. Zherdin, L. Tiemann, C. Plant, and M. Ploner, "Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data," *Cerebral Cortex*, vol. 22, no. 5, pp. 1118–1123, 2011.
[4] A. Williamson, and B. Hoggart, "Pain: A review of three commonly used pain rating scales," *J. Clin. Nursing*, vol. 14, no. 7, pp. 798–804, 2005.
[5] C. W. Woo *et al.*, "Quantifying cerebral contributions to pain beyond nociception," *Nature Commun.*, vol. 8, 2017, Art. no. 14211.
[6] M. C. Reddan and T. D. Wager, "Modeling pain using fMRI: From regions to biomarkers," *Neurosci. Bull.*, vol. 34, pp. 208–215, 2018.
[7] D. E. Harper, Y. Shah, E. Ichesco, G. E. Gerstner, and S. J. Peltier, "Multivariate classification of pain-evoked brain activity in temporomandibular disorder," *Pain Rep.*, vol. 1, no. 3, 2016, Art. no. e572.
[8] Y. Bai, G. Huang, Y. Tu, A. Tan, Y. S. Hung, and Z. Zhang, "Normalization of pain-evoked neural responses using spontaneous EEG improves the performance of EEG-based cross-individual pain prediction," *Front. Comput. Neurosci.*, vol. 10, no. 31, 2016.
[9] S. M. LaConte, S. J. Peltier, and X. P. Hu, "Real-time fMRI using brain-state classification," *Hum. Brain Mapping*, vol. 28, no. 10, pp. 1033–1044, 2007.
[10] Y. H. Tu *et al.*, "A novel and effective fMRI decoding approach based on sliced inverse regression and its application to pain prediction," *Neurocomputing*, vol. 273, pp. 373–384, 2018.
[11] L. Hu, and G. D. Iannetti, "Painful issues in pain prediction," *Trends Neurosci.*, vol. 39, no. 4, pp. 212–220, 2016.
[12] Q. Lin, "Influence of individual differences in fMRI-based pain prediction models on between-individual prediction performance," *Front. Neurosci.*, vol. 12, 2018, Art. no. 569.
[13] Y. Tu, A. Tan, Y. Bai, Y. S. Hung, and Z. Zhang, "Decoding subjective intensity of nociceptive pain from pre-stimulus and post-stimulus brain activities," *Front. Comput. Neurosci.*, vol. 10, no. 32, 2016, Art. no. 32.
[14] A. M. Anter and M. Ali, "Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems," *Soft Comput.*, vol. 24, pp. 1565–1584, 2019, doi: 10.1007/s00500-019-03988-3.
[15] M. Woźniak and D. Połap, "Bio-inspired methods modeled for respiratory disease detection from medical images," *Swarm Evol. Comput.*, vol. 41, pp. 69–96, 2018.

[16] A. M. Anter, Y. S. Moemen, A. Darwish, and A. E. Hassanien, "Multi-target QSAR modelling of chemo-genomic data analysis based on extreme learning machine," *Knowl.-Based Syst.*, vol. 188, 2019, Art. no. 104977.
[17] M. Woźniak and D. Połap, "Hybrid neuro-heuristic methodology for simulation and control of dynamic systems over time interval," *Neural Netw.*, vol. 93, pp. 45–56, 2017.
[18] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," *Comput. Struct.*, vol. 169, pp. 1–12, 2016.
[19] S. Hinojosa *et al.*, "Improving multi-criterion optimization with chaos: A novel multi-objective chaotic crow search algorithm," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 319–335, 2018.
[20] D. S. Tahir, and R. S. Ali, "A chaotic crow search algorithm for high-dimensional optimization problems," *Basrah J. Eng. Sci.*, vol. 17, no. 1, pp. 16–25, 2017.
[21] R. A. Poldrack, "Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding," *Neuron*, vol. 72, no. 5, pp. 692–697, 2011.
[22] P. Angelov, and R. Yager, "A new type of simplified fuzzy rule-based system," *Int. J. Gen. Syst.*, vol. 41, no. 2, pp. 163–185, 2012.
[23] A. M. Anter and A. E. Hassenian, "CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and adaptive watershed algorithm," *Artif. Intell. Med.*, vol. 97, pp. 105–117, 2019.
[24] T. P. Q. Nguyen and R. J. Kuo, "Automatic fuzzy clustering using non-dominated sorting particle swarm optimization algorithm for categorical data," *IEEE Access*, vol. 7, pp. 99721–99734, 2019.
[25] A. M. Anter, D. Gupta, and O. Castillo, "A novel parameter estimation in dynamic model via fuzzy swarm intelligence and chaos theory for faults in wastewater treatment plant," *Soft Comput.*, vol. 24, pp. 111–129, 2019, doi: 10.1007/s00500-019-04225-7.
[26] P. P. Angelov and X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1462–1475, Dec. 2008.
[27] X. Gu, and P. P. Angelov, "Self-organising fuzzy logic classifier," *Inf. Sci.*, vol. 447, pp. 36–51, 2018.
[28] A. M. Anter *et al.*, "A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI," *Inf. Sci.*, vol. 503, pp. 670–687, 2019.
[29] A. M. Anter, A. E. Hassenian, and D. Oliva, "An improved fast fuzzy c-means using crow search optimization algorithm for crop identification in agricultural," *Expert Syst. Appl.*, vol. 118, pp. 340–354, 2019.
[30] A. M. Anter and A. E. Hassenian, "Computational intelligence optimization approach based on particle swarm optimizer and neutrosophic set for abdominal CT liver tumor segmentation," *J. Comput. Sci.*, vol. 25, pp. 376–387, 2018.
[31] P. Angelov and X. Gu, "Deep rule-based classifier with human-level performance and characteristics," *Inf. Sci.*, vol. 463, pp. 196–213, 2018.
[32] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "DBSMOTE: Density-based synthetic minority over-sampling technique," *Appl. Intell.*, vol. 36, no. 3, pp. 664–684, 2012.
[33] J. F. Diez-Pastor, J. J. Rodríguez, C. Garcia-Osorio, and L. I. Kuncheva, "Random balance: Ensembles of variable priors classifiers for imbalanced data," *Knowl.-Based Syst.*, vol. 85, pp. 96–111, 2015.
[34] L. Y. Chuang, C. H. Yang, and J. C. Li, "Chaotic maps based on binary particle swarm optimization for feature selection," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 239–248, 2011.
[35] P. Angelov and R. Buswell, "Identification of evolving fuzzy rule-based models," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 5, pp. 667–677, Oct. 2002.
[36] E. Lughofer and P. Angelov, "Handling drifts and shifts in on-line data streams with evolving fuzzy systems," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2057–2068, 2011.
[37] J. S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.
[38] T. Yao, V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.
[39] G. Bortolan and W. Pedrycz, "Reconstruction problem and information granularity," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 234–248, May 1997.
[40] SPM. (2019). Accessed: Jun. 2018. [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/
[41] J. E. Brown, N. Chatterjee, J. Younger, and S. Mackey, "Towards a physiology-based measure of pain: Patterns of human brain activity distinguish painful from non-painful thermal stimulation," *PloS One*, vol. 6, no. 9, 2011, Art. no. e24124.
[42] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.
[43] R. J. Frackowiak *et al.*, *Human Brain Function*, 2nd ed. New York, NY, USA: Elsevier, 2013, pp. 1–1144.
[44] H. Chen, H. Yuan, D. Yao, L. Chen, and W. Chen, "An integrated neighborhood correlation and hierarchical clustering approach of functional MRI," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 452–458, Mar. 2006.
[45] L. Zhang and P. N. Suganthan, "A comprehensive evaluation of random vector functional link networks," *Inf. Sci.*, vol. 367, pp. 1094–1105, 2016.