IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 1, 2020, accepted June 12, 2020, date of publication June 16, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002904

Fuzzy Granular Hyperplane Classifiers

WEI LI[1], ZHONGNAN WEI[1], YUMIN CHEN[1], (Member, IEEE), CHAO TANG[2], AND YUPING SONG[3]

[1]School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China
[2]School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China
[3]School of Mathematical Sciences, Xiamen University, Xiamen 361005, China

Corresponding author: Wei Li (drweili@hotmail.com)

This work was supported in part by the Science and Technology Planning Guidance Project of Xiamen of China under Grant 3502Z20179038, in part by the National Natural Science Foundation of China under Grant 61976183, in part by the Natural Science Foundation of Fujian Province of China under Grant 2019J01850, in part by the Key Teaching and Research Project of Hefei University of China under Grant 2018 hfjyxm09, and in part by the Natural Science Foundation of Anhui Province under Grant 2008085MF202.

ABSTRACT Granular computing has advantage of knowledge discovery for complex data. In the paper, we present Fuzzy Granular Hyperplane Classifiers (FGHCs) for data classification from a new angle of Granular Computing. First, we introduce a fuzzy granular hyperplane concept by defining fuzzy granule, fuzzy granular vector, metrics and operators. Next, for binary classification problem, we present solving optimal fuzzy granular hyperplane through evolution strategy; the learning algorithm of parameters and the prediction algorithm of instances are also proposed. Finally, a multi-classification prediction model is designed by combining a set of Fuzzy Granular Hyperplane Classifiers based on vote strategy. In order to evaluate performance, we employed 10-fold cross validation to verify on UCI dataset and Alzheimer's Disease Voice dataset. Theoretical analysis and experiments demonstrated that FGHCs have good performance.

INDEX TERMS Fuzzy granular hyperplane, machine learning, granular computing.

I. INTRODUCTION

The classification of data is a general task in artificial intelligence. Assuming that some data points belong to one of two categories respectively, it is a goal for a new data point to make a correct decision to classify. Although there are lots of hyperplanes that can classify data points, there must exist an optimal hyperplane that can divide two categories at maximum margin. In this paper, we will discuss how to get such the best hyperplane in granular space based on fuzzy granulation to implement data classification with high accuracy.

In machine learning, Support Vector Machines (SVMs, [1]) is one of supervised learning approaches that can analyze trained data for classification. It is also a machine learning approach on the basis of principle of statistical learning theory. It includes superiority in prediction of small-scale instance sets, high-dimensional and nonlinear pattern recognition problems, and it can largely avoid the problems of "dimensional disaster" and "over-fitting". Also, it has a solid theoretical foundation, a simple and straightforward mathematical model. Hence in the area of pattern recognition, time series prediction, function estimation, regression analysis,etc., it has made considerable progress. It is also widely used in EEG signal analysis [2], [3], software effort estimation [4], disease detection [5], [6], medical image recognition [7], molecular and materials application [8], quantum computing [9]–[11] and so on.

The standard SVMs learning algorithms may be summarized as solving a quadratic programming (QP) problem with constraints. For a small-scale quadratic optimization problem, classical algorithms such as Newton's method and interior point approach can get good solutions. However, when training set is large, the complexity of algorithms will be so high that the efficiency will be low. At present, some advanced training algorithms are to decompose a complex QP problem into a suite of small-scale QP problems. According to some iterative strategies, these small QP problems can be solved respectively. Then, the approximate solution of the original large-scale QP problem can be calculated. Moreover, it will be gradually converged to the optimal solution.

In recent years, some scholars have also proposed some new SVMs methods like granular SVMs. The main strategy of granular SVMs is as follows: first, a series of information

The associate editor coordinating the review of this manuscript and approving it for publication was Pengcheng Liu.

granules can be obtained by constructing the granular space by division; then, learning is required in each information granule; after that, a decision function on SVMs can be obtained by clustering information (data, rules or attributes etc.) of granules [12]–[14]. The learning principle can convert a linearly inseparable problem into a suite of linearly separable ones through data granulation to obtain some decision functions. This also intensifies generalization performance of model. In other words, it can be obtained a hyperplane with wider margin in the training processing.

To overcome influences of noise and outliers on the SVMs [15], [16], Lin and his colleagues [17] proposed a Fuzzy Support Vector Machines (FSVM) by combining fuzzy mathematics and SVMs, which has good performance in processing noise data. The primary idea of this approach is to add a membership degree in the training instance set; the membership degree of support vectors is much larger, but that of non support vectors and outliers are smaller; thus, impacts from noise points, outliers and non support vectors to an optimal hyperplane can be reduced by the membership degree. It is of importance for FSVM how to determine the membership value, i.e., the weight of each instance. Some researchers proposed a membership function determined method based on class center; they adopted the distance between a instance point and its cluster as the degree of membership [17]–[19]. In the case, the membership function may depend heavily on the geometry of instances, so it may reduce the degree of membership associated with support-vectors. Zhang *et al.* [20] presented an approach about membership degrees determined on the basis of intra class hyperplane distance; they employed the distance between a instance point and its inner hyperplane as the membership function. In their scheme, membership functions rely lightly on geometry of a instance set and support vectors can obtain a larger degree of membership.

Statistical theory systematically studies machine learning problems, especially in the case of limited instances. SVMs and relative algorithms generated under theoretical framework show many superior performances in theory and practical applications.

Granular computing is a discipline that specializes in thinking, problem solutions and theory of information processing patterns on the basis of granular structure. It is also a new computing paradigm in the study of intelligent information handling. From angle of AI, granular computing is a natural model that simulates human thinking and solves complex large-scale problems. Deriving from practical problems, it can replace an exact solution with an approximate solution to achieve the purpose of simplifying the problem and improving the solving efficiency.

As early as 1979, Zade presented fuzzy set theory and fuzzy information granulation problem at first [21]. He thought that human cognitive ability can be summed up three primary characteristics of granulation, organization and causality [22]–[25]. In 1997, Zadeh first proposed the concept of granular computing [23]. Then lots of scholars in the world studied the problem, and increasingly established a new direction in artificial intelligence. Pedrycz [26] identified the principles of Granular Computing and showed how granules are built and then adopted in giving description of data relationships. Yao [27] first presented three-way decisions concept, acceptance, abstain and reject, in 2009. After that, a summary of three-way decisions theory was built on the basis of the notions of acceptance, rejection, and noncommitment [28], [29]. The theory is an addition of binary-decision model. Miao and his colleagues [30] proposed three-level model of granular spaces (the universe, the basis and the granular structure) in set-theoretic formulation. They established three-level model of granular spaces in Pawlak rough sets [31] by using the definability defined by the logic language Wang *et al.* [32] presented the diagram for relationship between three basic modes of granular computing. Also, they analyzed the feasibility of granular computing for big data processing. Hu and his colleagues [33] measured quality of features in multi-label learning by introducing fuzzy mutual information and developed effective approaches to guide section of multi-label feature. The algorithms can select from streaming features and be used for ordinal multi-label learning.

In addition, Granular Computing shows many conceptual developments, such as graphs [34], information tables [35], knowledge representation [36], clustering [37], rule clustering [38], classification [39] etc. There are a lot of applications of Granular Computing. And these applications like forecasting time series [40], [41], manufacturing [42], search encryption voice [43], prediction tasks [44], concept learning [45], optimization [46], K-nearest granule classifiers [47], attribute reduction [48], analysis of microarray data [49] and so on, are reported in recent studies. It is of worthy of emphasizing that information granules infiltrate almost all researches.

In this paper, method proposed is composed of two phases, parameters learning and prediction. When parameters learning, we convert instances into fuzzy granules. In the fuzzy granular space, loss function is created and parameters are solved by evolution strategy. When predicting, instances are converted into fuzzy granules and predicted by decision function. The processes are almost handled in the fuzzy granular space. Overview is as shown in Figure 1.

Contributions of the paper have three aspects. First, We define concepts of a fuzzy granule, a fuzzy granular vector, metrics and operators. Moreover, the concept of a fuzzy granular hyperplane is introduced. Second, on the basis of these, we present quickly solving optimal fuzzy granular hyperplane parameters through the evolution strategy. According to them, we design algorithms on parameters' learning and instances' prediction, which can solve a binary classification problem. Third, for a multi-classification problem, we adopt divide-and-conquer strategy, i.e., transforming a multi-classification problem into several binary classification problems, to obtain solutions. Also, we design a multi-class prediction model based on FGHCs.

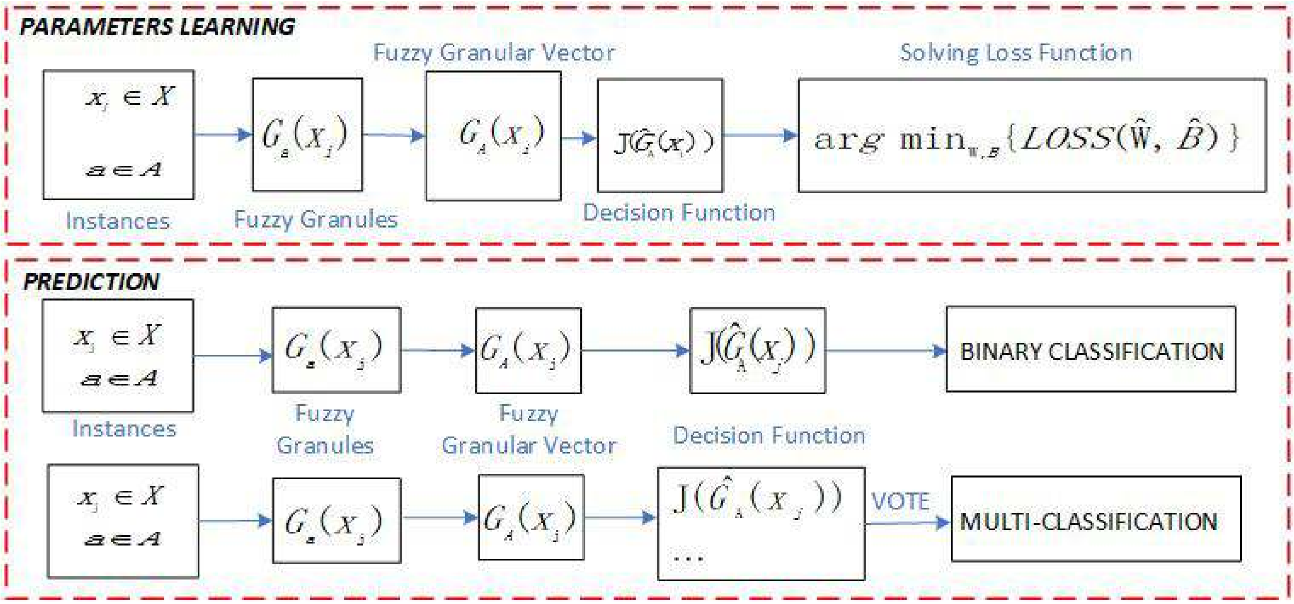


FIGURE 1. Overview of method.

II. FROM GRANULAR COMPUTING TO CLASSIFICATION

In lots of cases, granularity of human thinking and conceptual building are uncertain, rather than exact. Information fuzzy granulation is generally achieved by a fuzzy binary relationship, and fuzzy granulation is executed in whole fuzzy granular space. In this paper, fuzzy granules are obtained by fuzzy granulation, and then a fuzzy granular vector is formed by the granules, and a fuzzy granular hyperplane is constructed by fuzzy granular vectors. Thus, a fuzzy granular space is composed of fuzzy hyperplanes. In a fuzzy granular space, we also define measurements and operators, discuss Monotonic property and prove it. Based on these, we transform a classification problem of instances into hyperplane classification problem in fuzzy granular space. If the training data is linearly separable, optimal hyperplane that separate two categories of data can be selected, so that the distance between them is as large as possible. By solving the optimal fuzzy granular hyperplane with maximum margin, we can categorize unlabeled data.

A. DESIGN METHOD AND PRIMARY CHARACTERIZATION OF GRANULES

1) FROM INSTANCES TO FUZZY GRANULES

Definition 1: Let $D = (X, A, L)$ be a decision system, where $X = \{x_1, x_2, \ldots, x_n\}$ are n instances, $A = \{a_1, a_2, \ldots, a_m\}$ are m-dimensional features and $L = \{l\}$ $(l \in \{-1, +1\})$ is a label set, respectively.

Definition 2: For $\forall x_i, x_j \in X$ and $\forall a \in A$, a distance between x_i and x_j on feature a can be defined by:

$$d_a(x_i, x_j) = |h_a(x_i) - h_a(x_j)| \tag{1}$$

Here, $h_a(x_i)$ and $h_a(x_j)$ denote the normalization values of instance x_i and x_j on feature a, respectively. It's easy to have $d_a(x_i, x_j) \in [0, 1]$.

Definition 3: For $\forall x_i \in X, \forall a \in A$, the fuzzy granule of the instance x_i on feature a is defined as:

$$G_a(x_i) = d_{i1}/x_1 + d_{i2}/x_2 + \ldots + d_{in}/x_n \tag{2}$$

To simply the representation, d_{ij} is distance between x_i and x_j on feature a, i.e., $d_{ij} = d_a(x_i, x_j)$; "+" is union operator and "/" is separator. In that way, a fuzzy granule of instance is also a set which consists of pairs constructed by distance and instance.

Definition 4: For $\forall x \in X$ and $\forall a \in A$, the cardinality of fuzzy granule $G_a(x)$ can be quantified by:

$$|G_a(x)| = \sum_{t \in X} d_a(x, t) \tag{3}$$

We can easily get $0 \leq |G_a(x)| \leq |X|$ because of $d_a(x, t) \in [0, 1]$, where $|X|$ expresses the number of elements in the set X.

Definition 5: For $\forall s \in X$ and $Q \subseteq A$, assuming $Q = \{a_1, a_2, \ldots, a_k\}, (k \leq m)$, then the fuzzy granular vector of instance x on feature subset Q can be denoted by:

$$\hat{G}_Q(s) = (G_{a_1}(x), G_{a_2}(x), \ldots, G_{a_k}(x)) \tag{4}$$

Definition 6: For $\forall x \in X$ and $Q \subseteq A$, let $Q = \{a_1, a_2, \ldots, a_k\}, (k \leq m)$, then we can define the module of fuzzy granular vector of x on Q as follows:

$$|\hat{G}_Q(x)| = \sum_{a \in Q} |G_a(x)| \tag{5}$$

Definition 7: Let $G_a(x)$ and $G_a(t)$ be fuzzy granules of x and t on feature a, respectively, then in fuzzy granular space, operators $\cap$, $\cup$, $\sim$,and $\oplus$ can be defined as follows:

$$d_{min,i} = d_a(x_i, x) * d_a(x_i, t) \tag{6}$$

$$d_{max,i} = d_a(x_i, x) + d_a(x_i, t) - d_a(x_i, x) * d_a(x_i, t) \tag{7}$$

$$G_a(x) \cap G_a(t) = d_{min,1}/x_1 + d_{min,2}/x_2 + \ldots + d_{min,n}/x_n \tag{8}$$

$$G_a(x) \cup G_a(t) = d_{max,1}/x_1 + d_{max,2}/x_2 + \ldots + d_{max,n}/x_n \tag{9}$$

$$\sim G_a(x) = (1 - d_{i1})/x_1 + (1 - d_{i2})/x_2 + \ldots + (1 - d_{in})/x_n \tag{10}$$

$$G_a(x) \oplus G_a(t) = (d_{max,1} - d_{min,1})/x_1 + (d_{max,2} - d_{min,2})/x_2 + \ldots + (d_{max,n} - d_{min,n})/x_n \tag{11}$$

To be simplified, here $d_{ij} = d_a(x_i, x_j)$; "+" expresses union operator; "/" denotes separator between distance and instance.

Theorem 1: Let $G_a(x)$ and $G_a(t)$ be two fuzzy granules on a, formula (6) and (7), that is $d_{min,i}$ and $d_{max,i}$, have the following equation established:

$$0 < d_{min,i} < d_{max,i} \le 1 \tag{12}$$

Proof: Because of $0 < d_a(x_i, x) < 1$, $0 < d_a(x_i, t) < 1$, we have $d_a(x_i, x) * d_a(x_i, t) < d_a(x_i, x)$ and $d_a(x_i, x) * d_a(x_i, t) < d_a(x_i, t)$. So we get $2 * d_a(x_i, x) * d_a(x_i, t) < d_a(x_i, x) + d_a(x_i, t)$. Then, we also have $d_a(x_i, x) * d_a(x_i, t) < d_a(x_i, x) + d_a(x_i, t) - d_a(x_i, x) * d_a(x_i, t)$. Therefore, according to Equation (6) and (7), the formal $0 < d_{min,i} < d_{max,i}$ is established. Meanwhile, $d_a(x_i, x) + d_a(x_i, t) - d_a(x_i, x) * d_a(x_i, t) = d_a(x_i, x)(1 - d_a(x_i, t)) + d_a(x_i, t) < (1 - d_a(x_i, t)) + d_a(x_i, t) = 1$. That is, the formal $d_{max,i} \le 1$ is established. In sum, $0 < d_{min,i} < d_{max,i} \le 1$ is established. □

Definition 8: For $\forall x, t \in X$, let $\hat{G}_A(x) = (G_{a_1}(x), G_{a_2}(x), \ldots, G_{a_m}(x))$ and $\hat{G}_A(t) = (G_{a_1}(t), G_{a_2}(t), \ldots, G_{a_m}(t))$ be two fuzzy granular vectors on feature set A respectively, then operators $\cap$, $\cup$, $\sim$, and $\oplus$ can be defined as follows:

$$\hat{G}_A(x) \cap \hat{G}_A(t) = (G_{a_1}(x) \cap G_{a_1}(t), G_{a_2}(x) \cap G_{a_2}(t), \ldots, G_{a_m}(x) \cap G_{a_m}(t)) \tag{13}$$

$$\hat{G}_A(x) \cup \hat{G}_A(t) = (G_{a_1}(x) \cup G_{a_1}(t), G_{a_2}(x) \cup G_{a_2}(t), \ldots, G_{a_m}(x) \cup G_{a_m}(t)) \tag{14}$$

$$\sim \hat{G}_A(x) = (\sim G_{a_1}(x), \sim G_{a_2}(x), \ldots, \sim G_{a_m}(x)) \tag{15}$$

$$\hat{G}_A(x) \oplus \hat{G}_A(t) = (G_{a_1}(x) \oplus G_{a_1}(t), G_{a_2}(x) \oplus G_{a_2}(t), \ldots, G_{a_m}(x) \oplus G_{a_m}(t)) \tag{16}$$

Definition 9: For $\forall x, t \in X$, let $\hat{G}_A(x) = (G_{a_1}(x), G_{a_2}(x), \ldots, G_{a_m}(x))$ and $\hat{G}_A(t) = (G_{a_1}(t), G_{a_2}(t), \ldots, G_{a_m}(t))$ be two fuzzy granular vectors on feature set A respectively, then their distance can be defined as follows:

$$\hat{d}(\hat{G}_A(x), \hat{G}_A(t)) = \frac{1}{|A| * |X|} \sum_{a \in A} \frac{|G_a(x) \oplus G_a(t)|}{|G_a(x) \cup G_a(t)|} \tag{17}$$

TABLE 1. A decision system.

X	a	b	c	L
x_1	0.1	0.3	0.4	-1
x_2	0.3	0.2	0.3	-1
x_3	0.4	0.1	0.2	+1
x_4	0.2	0.4	0.1	+1

Theorem 2: For $\forall s, t \in X$, the distance of fuzzy granular vector satisfies:

$$0 < \hat{d}(\hat{G}_A(s), \hat{G}_A(t)) \le 1 \tag{18}$$

Proof: Assuming that $s = x_i$, $t = x_j$, according definition 2 and 3, we have $G_a(x_i) = d_{i1}/x_1 + d_{i2}/x_2 + \ldots + d_{in}/x_n$, $G_a(x_j) = d_{j1}/x_1 + d_{j2}/x_2 + \ldots + d_{jn}/x_n$.
From definition 1, we can get the distance $d_{ij} = d_a(x_i, x_j) \in (0, 1]$.

Then, we also have that the cardinality $|G_a(s)| = \sum_{t \in X} d_a(s, t)$ from definition 4. As definition 5 mentioned above, we can obtain two fuzzy granular vectors, $\hat{G}_A(x_i) = (G_{a_1}(x_i), G_{a_2}(x_i), \ldots, G_{a_m}(x_i))$, and $\hat{G}_A(x_j) = (G_{a_1}(x_j), G_{a_2}(x_j), \ldots, G_{a_m}(x_j))$.

Furthermore, from equation (6)-(13), we can get $\forall a \in A$ $0 < \frac{|G_a(s) \oplus G_a(t)|}{|G_a(s) \cup G_a(t)|} \le |X|$.

Since $0 < \sum_{a \in A} \frac{|G_a(s) \oplus G_a(t)|}{|G_a(s) \cup G_a(t)|} \le |A| * |X|$ is established, we can have $0 < \frac{1}{|A| * |X|} \sum_{a \in A} \frac{|G_a(s) \oplus G_a(t)|}{|G_a(s) \cup G_a(t)|} \le 1$. According to equation (17), we have $\hat{d}(\hat{G}_A(s), \hat{G}_A(t)) = \frac{1}{|A| * |X|} \sum_{a \in A} \frac{|G_a(s) \oplus G_a(t)|}{|G_a(s) \cup G_a(t)|}$, that is, $0 < \hat{d}(\hat{G}_A(s), \hat{G}_A(t)) \le 1$ □

Theorem 3: (Monotonic) Let $D = (X, A, L)$ be a decision system, for $\forall x \in X$ and feature subset $P \subseteq F \subseteq A$, there exists two fuzzy granular vectors, $\hat{G}_P(x)$ and $\hat{G}_F(x)$ of x on P and F. They satisfy:

$$|\hat{G}_P(x)| \le |\hat{G}_F(x)| \tag{19}$$

Proof: From definition 5, we have that $\hat{G}_P(x) = (G_{a_1}(x), G_{a_2}(x), \ldots, G_{a_u}(x))$, $\hat{G}_F(x) = (G_{a_1}(x), G_{a_2}(x), \ldots, G_{a_v}(x))$. For $\forall a \in P$, its fuzzy granule is $G_a(x)$. Because of $P \subseteq F$, $a \in F$ is established. Its fuzzy granule satisfies $G_a(x) \in \hat{G}_F(x)$ and $|P| \le |F|$. Hence, $\sum_{a \in P} |G_a(x)| \le \sum_{a \in F} |G_a(x)|$, that is $|\hat{G}_P(x)| \le |\hat{G}_F(x)|$. □

The similarity measurement denotes similar degree between fuzzy granular vectors. We can adopt distance between fuzzy granular vectors to measure their similarity. We take an example as follows to explain it.

Example 1: Let $D = (X, A, L)$ be a decision system, as demonstrated in Table 1. Here, $X = \{x_1, x_2, x_3, x_4\}$ is a instance set, $A = \{a, b, c\}$ represents a feature set and $L = \{l\}$ denotes a label set.

A instance set $X = \{x_1, x_2, x_3, x_4\}$ can be fuzzy granulated on the feature a as follows. First, the distances among instances can be calculated as:

$d_a(x_1, x_1) = |0.1 - 0.1| = 0$, $d_a(x_1, x_2) = |0.1 - 0.3| = 0.2$, $d_a(x_1, x_3) = |0.1 - 0.4| = 0.3$, and $d_a(x_1, x_4) = |0.1 - 0.2| = 0.1$. On the basis of them, according to definition 3, because the fuzzy granule of instance x_1 on feature a is expressed formally as $G_a(x_1) = d_a(x_1, x_1)/x_1 + d_a(x_1, x_2)/x_2 + d_a(x_1, x_3)/x_3 + d_a(x_1, x_4)/x_4$, it can be denoted actually as $G_a(x_1) = 0/x_1 + 0.2/x_2 + 0.3/x_3 + 0.1/x_4$. Its cardinality is $|G_a(x_1)| = 0 + 0.2 + 0.3 + 0.1 = 0.6$. Similarly, the fuzzy granule on feature b is $G_b(x_1) = 0/x_1 + 0.1/x_2 + 0.2/x_3 + 0.1/x_4$ and its cardinality is $|G_b(x_1)| = 0 + 0.1 + 0.2 + 0.1 = 0.4$; the fuzzy granule on feature c is $G_c(x_1) = 0/x_1 + 0.1/x_2 + 0.2/x_3 + 0.3/x_4$ and its cardinality is $|G_c(x_1)| = 0 + 0.1 + 0.2 + 0.3 = 0.6$. Therefore, a fuzzy granular vector of instance x_1 on feature set A is $\hat{G}_A(x_1) = (G_a(x_1), G_b(x_1), G_c(x_1))$ and its module is $|\hat{G}_A(x_1)| = |G_a(x_1)| + |G_b(x_1)| + |G_c(x_1)| = 0.6 + 0.4 + 0.6 = 1.6$.

B. FUZZY GRANULAR HYPERPLANE CLASSIFICATION

1) PROBLEM PROPOSED

Let's review the problem first and give a solution. Assuming there are n linearly separable instances, there are numerous decision hyperplanes with zero error in the fuzzy granular space to separate them. Which decision hyperplane is optimal? For the decision hyperplane with zero error, we make the separable margin as large as possible. To this end, we transform the problem into a fuzzy granular space to solve. In a decision system, data is fuzzy granulated and converted into fuzzy granules divided by different atom features. And then, these fuzzy granules can form a fuzzy granular vector. Furthermore, we combine instances' labels with their fuzzy granular vectors as pairs, which are rules. Then the rules may form a rule library. Therefore, in the rule library, the classified problem may be converted into the problem of searching the optimal fuzzy granular vector $\hat{W}$ and $\hat{B}$ in equation (22). The details are as follows: first, data is granulated by definition 2-5; next, we convert the classification problem into solving parameters $\hat{W}$ and $\hat{B}$ in equation (20); finally, for the purpose, a loss function is designed, as equation (22) exhibited, and we can get the optimal solution by minimizing the loss function.

Before solving, we also need to give some assumptions as follows:

(1) Errors from the predicted values and the true values are consistent with the Gaussian distribution.

(2) There are linearly separable solutions in the problem. In other words, we can find $\hat{W}$ and $\hat{B}$ to get the fuzzy granular hyperplane with the maximum margin between positive instances and negative ones.

We adopt evolution method to obtain the solution. Specifically, error returned from a loss function is distributed randomly to $\hat{W}$ and $\hat{B}$ to reduce the loss function value in next computing. After sever iterations, the loss function value will be reduced repeatedly until convergence and the optimal solution is obtained. Some definitions can be given on the problem as follows:

Definition 10: Let $D = (X, A, L)$ be a decision system, where $X = \{x_1, x_2, \ldots, x_n\}$ expresses a instance set, $A = \{a_1, a_2, \ldots, a_m\}$ denotes feature set, and $L = \{+1, -1\}$ represents a label set. For $\forall x \in X$, $l_x \in L$, there exists a pair or a rule, $lb_A(x) =< \hat{G}_A(x), l_x >$, composed of a fuzzy granular vector and a label on A. According to these pairs, a rule library, $LB_A = \{lb_A(x)|\forall x \in X\}$, is built. A m-dimensional fuzzy granular hyperplane in fuzzy granular space is defined by:

$$\hat{Y} = \hat{W} \cap \hat{G}_A(x) \cup \hat{B} \tag{20}$$

Here, the operators, "$\cap$" and "$\cup$", are defined by Equation (13) and (14). We define a decision function $J : \mathbb{G} \rightarrow (-\infty, +\infty)$ as follows, where $\mathbb{G}$ represents all fuzzy granular vector sets.

$$J(\hat{G}_A(x)) = ln|\hat{W} \cap \hat{G}_A(x) \cup \hat{B}| + \alpha \tag{21}$$

Here $\hat{W}$, $\hat{B}$ and $\hat{Y}$ are fuzzy granular vectors and α is a positive number. To classify, we define a loss function as:

$$LOSS(\hat{W}, \hat{B}) = \sum_{x \in X} |J(\hat{G}_A(x)) - l_x| + \lambda * |\hat{W}| \tag{22}$$

where $\lambda * |\hat{W}|$ is a regularization item to avoid overfitting, λ is a small positive number and $l_x \in \{-1, +1\}$ is a label of instance x.

2) PRINCIPLE OF SOLVING PARAMETERS

Input: instance set X, test instance t and maximum iteration times $MaxIters$

Output: the optimal fuzzy granular vectors, W^* and B^*

1. Delete the instances of missing some feature values.

2. Normalize the instance to values in [0, 1] by Equation (23):

$$\forall x_i \in X, \forall a \in A,$$
$$h_a(x_i) \leftarrow \frac{h_a(x_i) - min_{x_j \in X}\{h_a(x_j)\}}{max_{x_j \in X}\{h_a(x_j)\} - min_{x_j \in X}\{h_a(x_j)\}} \tag{23}$$

3. For each instance $x \in X$, runs step 4-6.

4. instance x is granulated on atom feature $a_i \in A$, we can get $G_{a_i}(x)$.

5. Form a fuzzy granular vector $\hat{G}_A(x) = (G_{a_1}(x), G_{a_2}(x), \ldots, G_{a_m}(x))$ of x.

6. Get the label l_x of x from the decision system.

7. Build a rule library $lb_A(x) =< \hat{G}_A(x), l_x >$ for training instances; For a test instance t, it satisfies $l_t = null$, where $null$ denotes uncertainty.

8. Generate n pairs of solutions randomly as initialization solution on i^{th} iteration, namely $\{(\hat{W_1}(i), \hat{B_1}(i)), (\hat{W_2}(i), \hat{B_2}(i)), \ldots, (\hat{W_n}(i), \hat{B_n}(i)\}$.

9. According to step 8, we can calculate the values of function $LOSS$ corresponding to each candidate solution, and they are sorted in ascending order of the value of function $LOSS$; the results are assigned to variable T, i.e., $T \leftarrow AscendSort(LOSS)$

10. Take the first half solutions of the T to the variable O, i.e., $O \leftarrow Select(T, 0.5)$

11. Each element of these new solutions is probably corrected with errors. According to the assuming condition mentioned above, the error satisfies $e \sim N(\mu, \sigma^2)$. Here,

TABLE 2. **Algorithm of solving parameters.**

Input:	a instance set S and maximum iterative times $MaxIters$
Output:	the optimal fuzzy granular vectors $\hat{W}^*$ and $\hat{B}^*$
1	Normalize the values of all instances according to features as follow: $\forall x_i \in X, \forall a \in A, h_a(x_i) \leftarrow \frac{h_a(x_i) - min_{x_j \in X}\{h_a(x_j)\}}{max_{x_j \in X}\{h_a(x_j)\} - min_{x_j \in X}\{h_a(x_j)\}}$
2	FOR $i = 1$ to $instanceNum$ $\exists x_i \in X$ FOR $j = 1$ to m On feature $\exists a_j \in A$, instance x is fuzzy granulated as $G_{a_j}(x_i)$; END FOR Form a fuzzy granular vector $\hat{G}_A(x_i) = (G_{a_1}(x_i), G_{a_2}(x_i), ..., G_{a_m}(x_i))$; Get the category of x_i, l_{x_i}; If x_i is a training instance, we can build a rule $lb_A(x_i) = < \hat{G}_A(x_i), l_{x_i} >$; If x_i is a test instance, $l_{x_i} = null$ ($null$ represents uncertain). END FOR

3	FOR iters=1 to MaxIters
	$WB(iters) \leftarrow \{(\hat{W}_1, \hat{B}_1), (\hat{W}_2, \hat{B}_2), ..., (\hat{W}_k, \hat{B_k})\}$;//Generate k pairs solutions randomly;
	$E_i \leftarrow LOSS(\hat{W}_i, \hat{B}_i), i = 1, 2, ..., k$.//Calculate loss function;
	$E \leftarrow \{E_1, E_2, ..., E_k\}$;
	$T \leftarrow SortAscend(E)$;//Sort by ascending order of errors and assign them to T
	$O \leftarrow Select(T, 1/2)$;//Select the first half of solutions of T and insert them into O.
	Create a pseudo-random number $r \in [\mu - 3\sigma, \mu + 3\sigma]$.
	We have a operation of plus or minus with probability P to each element of
	fuzzy granular vectors in O and the object is r.
	$WB^{'}(iters) \leftarrow \{(\hat{W_1'}, \hat{B_1'}), (\hat{W_2'}, \hat{B_2'}), ..., (\hat{W_{k/2}'}, \hat{B_{k/2}'})\}$;//Get new candidate solutions:
	$WB(iters) \leftarrow Select(WB \cup WB^{'}, k)$;// Combine old candidate solutions with new ones,
	//and select the best k pairs of candidate solutions as initialization solutions of
	//next iteration.
	END FOR

4	Return the optimal $\hat{W}^*$ and $\hat{B}^*$ that minimize the function $LOSS$.

e represents the error of each element of a fuzzy granular vector, μ denotes the mean and σ expresses standard deviation. First, we choose plus operation or minus one with probability P. Second, the error e is from a pseudo-random number, which satisfies $e \in [\mu - 3\sigma^2, \mu + 3\sigma^2]$, so we generate the error e in the range $[\mu - 3\sigma^2, \mu + 3\sigma^2]$ randomly and correct the candidate solutions. Thus, we can get some new ones.

12. Compare the new solutions with old ones according to equation (22) and take the better parts as the initial solutions for the next iteration.

13. Determine whether the number of iterations meets the request, if not, return to step 8; otherwise, return to step 14.

14. Return $\hat{W}^*$ and $\hat{B}^*$ that minimize the function $LOSS(\cdot)$.

3) ALGORITHM OF SOLVING PARAMETERS

According to the principle of solving the binary classification problem, we design an algorithm of solving parameters (see Table 2). For a multi classification problem, we need to divide it into a set of binary classification problems. By solving every binary classification problem, multi classification one can get solved.

4) PRINCIPLE OF BINARY CLASSIFICATION PREDICTION

In prediction stage, given a instance $x \in X$, we have a fuzzy granular vector $\hat{G}_A(x)$ by fuzzy information granulation. After solving parameters, we also get the optimal solution $\hat{W}^*$ and $\hat{B}^*$. The prediction value of instance x can be calculated by Equation (24).

$$l_x = \begin{cases} -1, & if\ J(\hat{G}_A(x), \hat{W}, \hat{B}) < 0 \\ +1, & if\ J(\hat{G}_A(x), \hat{W}, \hat{B}) \geq 0 \end{cases} \tag{24}$$

5) PRINCIPLE OF MULTI-CLASSIFICATION PREDICTION

We have introduced how to solve two-category problem as mentioned above. For a multi-classification problem, we can divide the multi-classification problem into a series of binary classification ones in advance. Then, we can solve each binary classification problem and employ voting approach to give final decision. For any binary classification instances, we design a fuzzy granular hyperplane classifiers (FGHCs). Then, given N-classification instance set, to solve the multi-classification prediction problem, we need to train $\frac{N(N-1)}{2}$ FGHCs. When predicting a test instance, we adopt each FGHCs to have a prediction result and count the votes. The category with most votes is the final prediction value of the test instance. We take an example to explain as follows: considering a classification task with three categories C_1, C_2, C_3, we choose the fuzzy granular vectors associated with $(C_1, C_2), (C_1, C_3), (C_2, C_3)$ respectively as three training sets. After training, we have three pairs optimal parameters. Given a test instance, we employ the three pairs parameters to predict the label and get three results. Moreover, we count these results and give the final decision by voting. In other words, the category with most votes will

TABLE 3. Algorithm of prediction on binary classification.

Input: instance x, fuzzy granular vector $\hat{W}^*$ and $\hat{B}^*$.
Output: Category l_x of instance x
1. Get fuzzy granular vector $\hat{G}_A(x)$ on instance x from fuzzy granulation;
2. Calculate decision function: $J(\hat{G}_A(x), \hat{W}^*, \hat{B}^*) = ln\|\hat{W}^* \cap \hat{G}_A(x) \cup \hat{B}^*\| + \alpha$
3. Obtain approximate value $l_x = \begin{cases} -1, & if \quad J(\hat{G}_A(x), \hat{W}^*, \hat{B}^*) < 0 \\ +1, & if \quad J(\hat{G}_A(x), \hat{W}^*, \hat{B}^*) \geq 0 \end{cases}$
4. Return l_x.

TABLE 4. Algorithm of prediction on multi-classification.

Input:	instance$x$; Number of Category N, Fuzzy Granule Vectors $(\hat{W}_1, \hat{B}_1), (\hat{W}_2, \hat{B}_2), ..., (\hat{W}_{N*(N-1)/2}, \hat{B}_{N*(N-1)/2})$.
Output:	Category l_x of instance x.
1	Get fuzzy granular vector $\hat{G}_A(x)$ on instance x from fuzzy granulation;
2	Initialize votes $Vote(C_i) = 0, i = 1, 2, ..., N$, where C_i denotes a category associated with a instance.
3	For $k = 1$ to $N * (N-1)/2$ Find out the relative parameters $(\hat{W}_k, \hat{B}_k) = FindParameter(k)$. Get category $(C_i \ or \ C_j) = FindClass(k), i, j = 1, 2, ..., N$ Calculate decision function: $J(\hat{G}_A(x), \hat{W}, \hat{B}) = ln\|\hat{W} \cap \hat{G}_A(x) \cup \hat{B}\| + \alpha$. $if J(\hat{G}_A(x), \hat{W}_k, \hat{B}_k) < 0$, then $Vote(C_i) = Vote(C_i) + 1$; else $Vote(C_j) = Vote(C_j) + 1$; End For
4	Count the votes of categories and the category with the most votes is final decision, i.e., $l_x = Argmax(Vote(C_j)), j = 1, 2, ..., N$.
5	Return l_x.

TABLE 5. Data set.

Data Set	Number of instances	Number of Features	Number of Category
banknote	1372	4	2
Iris	150	4	3
Seeds	210	7	3
Nomao	34465	120	2
Voice of Alzheimer's Disease	100,000	114	2

be the final decision. The processes are as follows: First, initialize counter $Vote(C_1) = Vote(C_2) = Vote(C_3) = 0$ of categories; Second, use FGHCs associated with (C_1, C_2) to predict, if the result is C_1, then $Vote(C_1) = Vote(C_1) + 1$, otherwise, $Vote(C_2) = Vote(C_2) + 1$; similarly, employ FGHCs of (C_1, C_3) to determine, if label predicted is C_1, then $Vote(C_1) = Vote(C_1) + 1$; if not, $Vote(C_3) = Vote(C_3) + 1$; adopt FGHC of (C_2, C_3) to give the prediction result, if it is C_2, then $Vote(C_2) = Vote(C_2) + 1$, or else $Vote(C_3) = Vote(C_3)+1$; Finally, the category with most votes will be the final result, that is, $Argmax(Vote(C_1), Vote(C_2), Vote(C_3))$.

6) ALGORITHM OF PREDICTION

After the parameter solution is executed, the fuzzy granular vector of test instances and optimal parameters of FGHCs can be obtained. We also design algorithms of prediction on binary classification and multi-classification based on them (see Table 3 and Table 4).

III. EXPERIMENTAL ANALYSIS

To measure how well FGHCs performed at classification problem, we evaluated the performance of the classifier using four datasets in UCI and one Alzheimer's Disease voice dataset, as shown in Table 5. 10-fold cross validation was adopted in the experimental results. The values of the dataset are various, so these values need to be normalized. The equation (23) was employed to ensure that all values can be normalized in [0, 1]. The data is fuzzy granulated on every atomic feature to build a fuzzy granule. Then, a fuzzy granular vector consists of these fuzzy granules. To verify the performance, we compared Back Propagation (BP), SVMs and FGHCs on evaluation indexes. We adopted True Positive (TP) rate, False Positive (FP) rate, Precision, Recall, F-score and ROC Area to evaluate the performance. In the evaluation, we exhibited the parameters of each category of instance and compare them with three algorithms as mentioned above. In running FGHCs, the max iteration times are 1000. Parameters of SVMs include penalty coefficient C and γ which can be regarded as the inverse of the radius of influence of instances selected by the model as support vectors. Parameters of BP involve the number of hidden layer N_h, the number of unit of input layer I_n, the number of unit of hidden layer H_n, and the number of output layer O_n, learning rate η, and maximum iterations M_n. Parameters of FGHCs involve maximum iterations M_I, the number of initialization solution k, adjust factor α, and penalization factor λ.

TABLE 6. Data set bank.

Method	TP Rate	FP Rate	Precision	Recall	F-score	ROC Area	Category
BP	0.990	0.009	0.989	0.990	0.989	1.000	1
	0.991	0.010	0.992	0.991	0.991	1.000	-1
SVMs	1.000	0.035	0.958	1.000	0.978	0.982	1
	0.965	0.000	1.000	0.965	0.982	0.982	-1
FGHCs	0.998	0.003	0.997	0.998	0.998	1.000	1
	0.997	0.002	0.999	0.997	0.998	1.000	-1

TABLE 7. Data set Iris.

Method	TP Rate	FP Rate	Precision	Recall	F-score	ROC Area	Category
BP	1.000	0.000	1.000	1.000	1.000	1.000	1
	0.920	0.020	0.958	0.920	0.939	0.972	2
	0.960	0.040	0.923	0.960	0.941	0.972	3
SVMs	1.000	0.000	1.000	1.000	1.000	1.000	1
	0.980	0.050	0.907	0.980	0.942	0.965	2
	0.900	0.010	0.978	0.900	0.938	0.970	3
FGHCs	1.000	0.000	1.000	1.000	1.000	1.000	1
	0.960	0.020	0.960	0.960	0.960	0.996	2
	0.960	0.020	0.960	0.960	0.960	0.996	3

As shown in Table 6, in the Bank dataset, parameters of BP were as follows: $N_h = 1$, $I_n = 4$, $H_n = 6$, $O_n = 2$, $\eta = 0.016$, and $M_n = 10000$. Parameters of SVMs satisfied to $C = 1$ and $\gamma = 0.1$. Parameters of FGHCs were as follows: $M_I = 10000$, $k = 400$, $\alpha = 0.217$, and $\lambda = 0.026$. FGHCs and BP achieved ROC Area of 1 in both categories, while SVMs got 0.982 (i.e., 0.81% improvement). F-score improved by 0.91% and 0.71% in two categories respectively compared with BP; it increased by 2.04% and 1.63% respectively compared with SVMs, exhibiting the quality of the positive predictions. FGHCs outperform BP in Recall metric over this dataset. Recall increased by 3.32% compared with SVMs in category -1, while decreased by 0.2% in category 1. Large improvements (i.e., 4.07%) in the Precision in class -1 compared with SVMs; meanwhile, Precision decreased by 0.2%. The Precisions of FGHCs are superior to those obtained by BP. Overall, FGHCs performs slightly better SVMs and BP in the Bank dataset.

In Iris dataset, there are three categories, as exhibited in Table 7. Parameters of BP were as follows: $N_h = 1$, $I_n = 4$, $H_n = 200$, $O_n = 3$, $\eta = 0.028$, and $M_n = 1000$. Parameters of SVMs satisfied to $C = 10$ and $\gamma = 0.01$. Parameters of FGHCs were as follows: $M_I = 1000$, $k = 100$, $\alpha = 0.128$, and $\lambda = 0.015$. ROC Area improved by 2.47% in category 2 and 3 compared with BP. It increased by 3.21% and 2.68% in two categories compared with SVMs. Interestingly, in category 1, all perform metrics of three methods are the same. F-score made an improvement by 2.24% and 2.02% in category 2 and 3, respectively, compared with BP. It decreased by 1.91% and 2.35% in the two categories compared with SVMs. Recall increased by 4.35% in class 2 compared with BP. Meanwhile, it improved by 6.67% in class 3, while it decreased by 2.08% in class 2; therefore, average increased by 1.39%, compared with SVMs. From the average of precision to evaluate, FGHCs was 0.9733, BP was 0.9603, and SVMs was 0.9617. The average of precision improved by 1.35% and 1.21%, respectively.

The number of dataset Seeds is between Bank's and Iris'. Parameters of BP were as follows: $N_h = 1$, $I_n = 7$, $H_n = 230$, $O_n = 3$, $\eta = 0.015$, and $M_n = 1000$. Parameters of SVMs satisfied to $C = 12$ and $\gamma = 0.021$. Parameters of FGHCs were as follows: $M_I = 1000$, $k = 120$, $\alpha = 0.113$, and $\lambda = 0.019$. Table 8 compared the performance metrics with the three approaches. The details are as follows. ROC Area of FGHCs was superior to those obtained by BP and SVMs. In particular, FGHCs achieved average ROC Area of more than 0.996, while BP was 0.980 and SVMs just got 0.964 (i.e., 1.63% and 3.32% improvement respectively). In F-score, the average was 0.952 from FGHCs, which improved by 1.49% compared with SVMs; it was the same as BP's. Recall of FGHCs in average was 0.953, BP's was 0.952 and SVMs' was 0.938. It increased by 0.11% and 1.60% respectively. The precision of FGHCs was also better than that of SVMs. It made an improvement by 1.6% in average and had almost not changed compared with BP's.

In dataset Nomao, it includes 2 categories, 120 attributes and 34465 instances. Parameters of BP were as follows: $N_h = 2$, $I_n = 120$, $H_{n_1} = 200$, $H_{n_2} = 230$ $O_n = 2$, $\eta = 0.022$, and $M_n = 200000$. Parameters of SVMs satisfied to $C = 7$ and $\gamma = 0.015$. Parameters of FGHCs were as follows: $M_I = 500000$, $k = 10000$, $\alpha = 0.287$, and $\lambda = 0.025$. As shown in Figure 2, the performance of FGHCs are superior to BP and SVMs. In particular, FGHCs achieved recall of 0.975 in category -1, while BP and SVMs just got 0.928 and 0.967 (i.e., 4.91% and 0.83% improvement respectively). Accuracy improved by 3.64% and 0.73% respectively. Recall

TABLE 8. Data set seeds.

Method	TP Rate	FP Rate	Precision	Recall	F-score	ROC Area	Category
BP	0.957	0.036	0.931	0.957	0.944	0.974	1
	0.986	0.014	0.972	0.986	0.979	0.996	2
	0.914	0.021	0.955	0.914	0.934	0.971	3
SVMs	0.900	0.043	0.913	0.900	0.906	0.929	1
	0.971	0.021	0.958	0.971	0.965	0.986	2
	0.943	0.029	0.943	0.943	0.943	0.978	3
FGHCs	0.929	0.036	0.929	0.929	0.929	0.993	1
	1.000	0.021	0.959	1.000	0.979	0.999	2
	0.929	0.014	0.970	0.929	0.949	0.996	3

of category +1 increased by 4.55% and 0.31% respectively. To verify further performance, we adopted voice dataset from University of Pittsburgh on Alzheimer's disease. The dataset is composed of voices and transcripts from participants. The corpus includes 1264 instances divided into two categories. To extend scale of the dataset, we splitted these voices into 100,000 fragments as a new dataset. As well known, Mel Cepstral Coefficients (MFCC) and Gammatone Cepstral Coefficients (GFCC) are both good features for classification. Here we extracted the two types of features of voice for classification. For MFCC, the first 20 dimensions, its first-order difference, and its second-order difference were selected and connected to obtain 57-dimensional features. Similarly, we extracted 57-dimensional features of GFCC. Thus we can get 114-dimensional features of Alzheimer's disease voice. The classification includes two classes. We compared long short-term memory (LSTM) with FGHCs on ROC as shown in Figure 3. Parameters of LSTM included 2 hidden layers with 200 hidden units at each layer. The learning rate belonged to $[0.0001, 0.01]$ and the maximum iterations was 1,000,000. Adam optimizer was used to do gradient optimization. Parameters of BP were as follows: $N_h = 2$, $I_n = 114$, $H_{n_1} = 100$, $H_{n_2} = 110$ $O_n = 2$, $\eta \in [0.01, 0.001]$, and $M_n = 1,000,000$. Parameters of SVMs satisfied to $C \in [18, 25]$ and $\gamma \in [0.001, 0.01]$. Parameters of FGHCs were as follows: $M_I = 1,000,000$, $k = 12,000$, $\alpha \in [0.001, 0.5]$, and $\lambda \in [0.0001, 0.05]$. Figure 3(a) and (b) compares ROC Area got by FGHCs, LSTM, SVMs, and BP using MFCC and GFCC feature. As demonstrated in Figure 3(a) and Figure 3(b), the ROC curve of FGHCs are closer to the upper left corner than other classifiers. Hence FGHCs had better performance than other models using MFCC and GFCC on the dataset. As shown in Figure 3(a), FGHCs got ROC Area of 0.92, and LSTM achieved ROC Area of 0.90 (i.e., improvement 2.22%). ROC Area of SVMs was 0.82 and ROC Area of BP was 0.85. FGHCs increased by 12.20% and 8.24% respectively. As shown in Figure 3(b), FGHCs improved by 2.76%, 13.02%, and 8.98% compared with LSTM, SVMs, and BP respectively.

In addition, we also employed sampling raw data as features to train these classifiers and compared ROC Area obtained by the classifiers (see Figure 3(c) and (d)). Parameters of LSTM involved 2 hidden layers with 210 hidden units

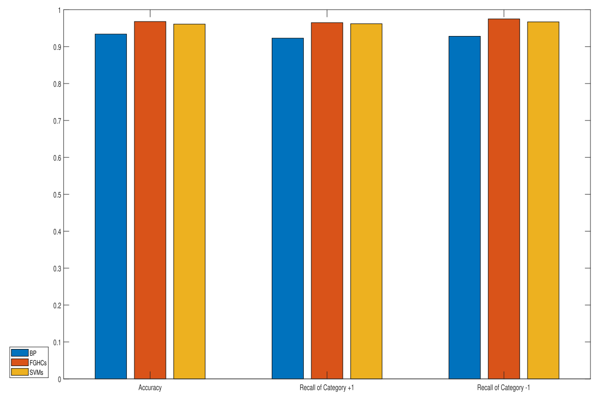


FIGURE 2. Comparison of accuracy and recall on data set Nomao.

at each layer. The learning rate belonged to $[0.0001, 0.01]$ and the maximum iterations was 1,000,000. Adam optimizer was also adopted to do gradient optimization. Parameters of BP were as follows: $N_h = 2$, $I_n = 180$, $H_{n_1} = 230$, $H_{n_2} = 240$, $O_n = 2$, $\eta \in [0.001, 0.0015]$, and $M_n = 1,000,000$. Parameters of SVMs satisfied to $C \in [15, 30]$ and $\gamma \in [0.001, 0.015]$. Parameters of FGHCs were as follows: $M_I = 1,000,000$, $k = 12,050$, $\alpha \in [0.001, 0.6]$, and $\lambda \in [0.0001, 0.09]$.

As shown in Figure 3(c), FGHCs got ROC Area of 0.912, LSTM achieved ROC Area of 0.903. FGHCs increased by 1.00%. Compared with SVMs and BP, FGHCs improved by 12.18% and 9.75% (SVMs was ROC Area of 0.813 and BP was ROC Area of 0.831). As demonstrated in Figure 3(d), SVMs got ROC Area of 0.823 and BP achieved ROC Area of 0.834. FGHCs obtained ROC Area of 0.887 and LSTM got ROC Area of 0.881. FGHCs increased by 7.78% and 6.35% compared with SVMs and BP respectively. FGHCs improved by 0.68% compared with LSTM.

Overall, FGHCs is superior to LSTM, SVMs, and BP. We also found that MFCC and GFCC features can lead to the improvement of performance.

In sum, FGHCs is superior to BP and SVMs as whole. FGHCs performs slightly better than LSTM. The main reasons are as follows. a) Fuzzy granulation is considered before classification and it embodies the angle of the collective structures of instances. b) While solving these optimal parameters,

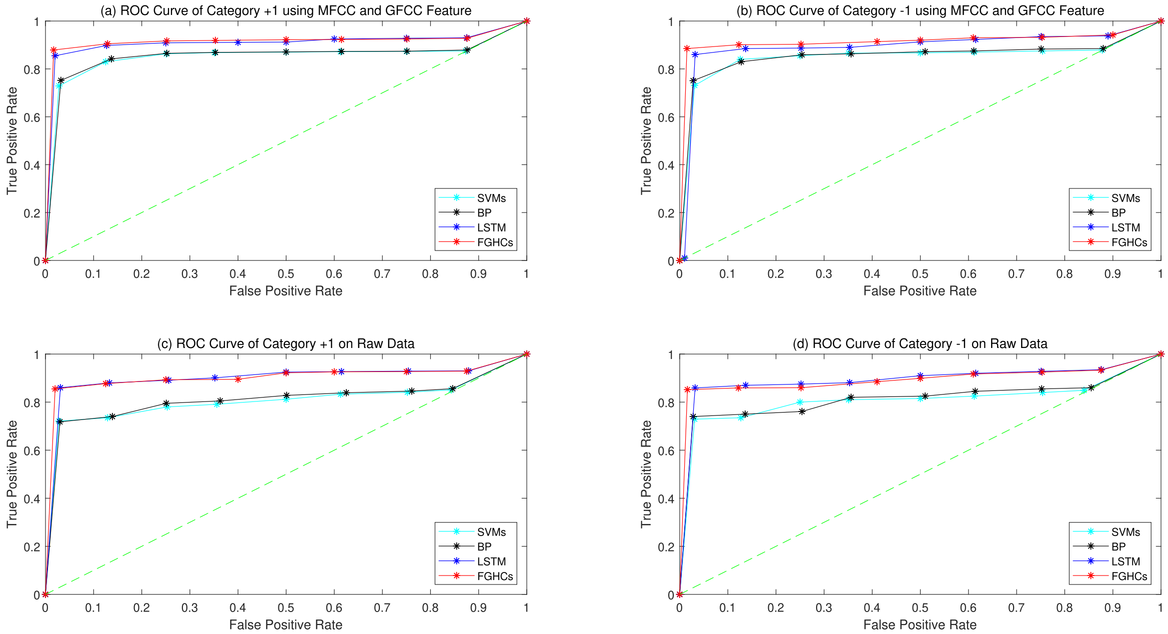


FIGURE 3. Comparison of ROC on data set of Alzheimer's disease voice.

we take into account the evolution principle that can find the global optimal solution. In contrast, BP and LSTM sometimes get the optimal solution that may be localized rather than global one; SVMs can obtain global optimal solution according to solving principle, but it just classifies the raw data instead of granules.

IV. CONCLUSION

In this paper, we propose Fuzzy Granular Hyperplane Classifiers from granular computing view. The scheme is as follows. First, we introduce a fuzzy granular hyperplane concept by some new definitions on fuzzy granules, operators and metrics. Then, we present parameters learning and instances predicted algorithms of binary classification based on evolution computing. To find multi-classification solution, we propose a predicted model by voting on the basis of a series of Fuzzy Granular Hyperplane Classifiers of binary classification. That is, we transform a multi-classification problem into a set of binary classification problems and employ counting votes to achieve the final decision. Experimental results show the performance of FGHCs outperforms those of BP, SVMs and LSTM under special parameters. In future work, we plan to add localized information granulation, parallel and distributed thoughts into algorithms to apply widely in big data research.

REFERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[2] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1106–1117, Feb. 2015.

[3] R. Varatharajan, G. Manogaran, and M. K. Priyan, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Multimedia Tools Appl.*, vol. 77, no. 8, pp. 10195–10215, 2018.

[4] K. Dejaeger, W. Verbeke, D. Martens, and B. Baesens, "Data mining techniques for software effort estimation: A comparative study," *IEEE Trans. Softw. Eng.*, vol. 38, no. 2, pp. 375–397, Mar. 2012.

[5] H. Akbari, L. Halig, M. D. Schuster, B. Fei, A. Osunkoya, V. Master, P. Nieh, and G. Chen, "Hyperspectral imaging and quantitative analysis for prostate cancer detection," *J. Biomed. Opt.*, vol. 17, no. 7, pp. 1–10, 2012.

[6] G. Orrù, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neurosci. Biobehav. Rev.*, vol. 36, no. 4, pp. 1140–1152, Apr. 2012.

[7] H. Selvaraj, S. T. Selvi, D. Selvathi, and L. Gewali, "Brain MRI slices classification using least squares support vector machine," *Int. J. Intell. Comput. Med. Sci. Image Process.*, vol. 1, no. 1, pp. 21–33, Jan. 2007.

[8] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, pp. 547–555, Jul. 2018.

[9] P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum support vector machine for big data classification," *Phys. Rev. Lett.*, vol. 113, no. 13, pp. 1–5, Sep. 2014.

[10] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, "Quantum machine learning: A classical perspective," *Proc. Roy. Soc. A*, vol. 474, no. 2209, pp. 1–26, 2018.

[11] Z. Li, X. Liu, N. Xu, and J. Du, "Experimental realization of a quantum support vector machine," *Phys. Rev. Lett.*, vol. 114, no. 14, Apr. 2015, Art. no. 140504.

[12] Y. Tang, B. Jin, Y. Sun, and Y.-Q. Zhang, "Granular support vector machines for medical binary classification problems," in *Proc. IEEE Int. Geosci. Remote Sens. (IGARSS)*, Oct. 2004, pp. 73–78.

[13] W. Cheng, Y. Shi, and Y.-P. Zhang, "Three primary mothods of granular computing," *Comput. Technol. Develop.*, vol. 17, no. 3, pp. 91–94, 2007.

[14] D. Shifei, X. Li, Z. Hong, and Z. Liwen, "Research and progress of cluster algorithms based on granular computing," *Int. J. Digit. Content Technol. Appl.*, vol. 4, no. 5, pp. 96–104, Aug. 2010.

[15] X. Li and L. Shu, "Fuzzy theory based support vector machine classifier," in *Proc. 5th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Shandong, China, Oct. 2008, pp. 600–604.

[16] P.-Y. Hao, M.-S. Lin, and L.-B. Tsai, "A new support vector machine with fuzzy hyper-plane and its application to evaluate credit risk," in *Proc. 8th Int. Conf. Intell. Syst. Design Appl.*, Taiwan, China, Nov. 2008, pp. 83–88.

[17] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Netw.*, vol. 3, no. 2, pp. 464–471, 2002.

[18] H. P. Huang and Y. H. Liu, "Fuzzy support vector machines for pattern recognition and data mining," *Int. J. Fuzzy Syst.*, vol. 4, no. 3, pp. 826–835, 2002.

[19] X. Zhang, X.-L. Zhang, and G.-Y. Xu, "Determination and analysis of fuzzy membership for SVM," *J. Image Graph.*, vol. 11, no. 8, pp. 1188–1192, 2006.

[20] G.-X. Zhang, L. Fei, Z. Du, and S.-Y. Liu, "Fuzzy support vector machine based on cluster hyperplane," *Comput. Eng. Des.*, vol. 12, no. 29, pp. 3177–3178, 2008.

[21] L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, M. Gupta, R. Ragade, and R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, 1979, pp. 3–18.

[22] L. A. Zadeh, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.

[23] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997.

[24] L. A. Zadeh, "Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems," *Soft Comput. Fusion Found., Methodologies Appl.*, vol. 2, no. 1, pp. 23–25, Apr. 1998.

[25] L. A. Zadeh, "A new direction in AI: Toward a computational theory of perceptions," *AI Mag.*, vol. 22, no. 1, pp. 73–84, 2001.

[26] W. Pedrycz, "Granular computing for data analytics: A manifesto of human-centric computing," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 6, pp. 105–120, Aug. 2018.

[27] Y. Yao, "Three-way decision: An interpretation of rules in rough set theory," in *Proc. 4th Int. Conf. Rough Sets Knowl. Technol. (RSKT)*, vol. 5589, 2009, pp. 642–649.

[28] Y. Y. Yao, "An outline of a theory of three-way decisions," in *Rough Sets and Current Trends in Computing* (Lecture Notes in Computer Science), vol. 7413, Chengdu, China, 2012, pp. 1–17.

[29] Y. Yao, J. Qi, and L. Wei, "Formal concept analysis, rough set analysis and granular computing based on three-way decisions," *J. Northeast Univ.*, vol. 48, no. 4, pp. 477–487, 2018.

[30] D.-Q. Miao, F.-F. Xu, Y.-Y. Yao, and L. Wei, "Set-theoretic formulation of granular computing," *Chin. J. Comput.*, vol. 35, no. 2, pp. 351–363, Aug. 2012.

[31] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, p. 341–356, 1982.

[32] G. Wang, J. Yang, and J. Xu, "Granular computing: From granularity optimization to multi-granularity joint problem solving," *J. Northeast Univ.*, vol. 2, no. 3, p. 105–120, 2017.

[33] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, "Streaming feature selection for multilabel learning based on fuzzy mutual information," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1491–1507, Dec. 2017.

[34] Q. Wang and Z. Gong, "An application of fuzzy hypergraphs and hypergraphs in granular computing," *Inf. Sci.*, vol. 429, pp. 296–314, Mar. 2018.

[35] G. Chiaselotti, T. Gentile, and F. Infusino, "Granular computing on information tables: Families of subsets and operators," *Inf. Sci.*, vols. 442–443, pp. 72–102, May 2018.

[36] G. Chiaselotti, T. Gentile, and F. Infusino, "Knowledge pairing systems in granular computing," *Knowl.-Based Syst.*, vol. 124, pp. 144–163, May 2017.

[37] X.-Q. Tang and P. Zhu, "Hierarchical clustering problems and analysis of fuzzy proximity relation on granular space," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 814–824, Oct. 2013.

[38] X. Wang, X. Liu, and L. Zhang, "Granular flow graph, adaptive rule generation and tracking," *Appl. Soft Comput.*, vol. 24, pp. 534–542, Aug. 2014.

[39] A. V. Savchenko, "Fast multi-class recognition of piecewise regular objects based on sequential three-way decisions and granular computing," *Knowl.-Based Syst.*, vol. 91, pp. 252–262, Jan. 2016.

[40] P. Singh and G. Dhiman, "A hybrid fuzzy time series forecasting model based on granular computing and bio-inspired optimization approaches," *J. Comput. Sci.*, vol. 27, pp. 370–385, Jul. 2018.

[41] O. Hryniewicz and K. Kaczmarek, "Bayesian analysis of time series using granular computing approach," *Appl. Soft Comput.*, vol. 47, pp. 644–652, Oct. 2016.

[42] J. Leng, Q. Chen, N. Mao, and P. Jiang, "Combining granular computing technique with deep learning for service planning under social manufacturing contexts," *Knowl.-Based Syst.*, vol. 143, pp. 295–306, Mar. 2018.

[43] W. Li, Y. Chen, H. Hu, and C. Tang, "Using granule to search privacy preserving voice in home IoT systems," *IEEE Access*, vol. 8, pp. 31957–31969, 2020.

[44] Z. Han, J. Zhao, H. Leung, and W. Wang, "Construction of prediction intervals for gas flow systems in steel industry based on granular computing," *Control Eng. Pract.*, vol. 78, pp. 79–88, Sep. 2018.

[45] J. Li, C. Mei, W. Xu, and Y. Qian, "Concept learning via granular computing: A cognitive viewpoint," *Inf. Sci.*, vol. 298, pp. 447–467, Mar. 2015.

[46] J. Martínez-Frutos, P. J. Martínez-Castejón, and D. Herrero-Pérez, "Efficient topology optimization using GPU computing with multilevel granularity," *Adv. Eng. Softw.*, vol. 106, pp. 47–62, Apr. 2017.

[47] W. Li, Y. Chen, and Y. Song, "Boosted K-nearest neighbor classifiers based on fuzzy granules," *Knowl.-Based Syst.*, vol. 195, pp. 1–13, May 2020.

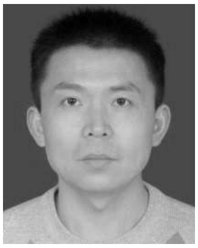
[48] K. Liu, X. Yang, H. Yu, H. Fujita, X. Chen, and D. Liu, "Supervised information granulation strategy for attribute reduction," *Int. J. Mach. Learn. Cybern.*, early access, Mar. 2020. [Online]. Available: https://link.springer.com/article/10.1007/s13042-020-01107-5, doi: 10.1007/s13042-020-01107-5.

[49] S. S. Ray, A. Ganivada, and S. K. Pal, "A granular self-organizing map for clustering and gene selection in microarray data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 9, pp. 1890–1906, Sep. 2016.

WEI LI received the Ph.D. degree from Xiamen University, China, in 2013. He was a Visiting Scholar with the Department of Computer Science, University of Massachusetts, Boston, USA, from June 2018 to June 2019. He is currently an Associate Professor with the School of Computer and Information Engineering, Xiamen University of Technology, China. His research interests include machine learning, fuzzy modeling and granular computing, big data, and information security. He has published over 20 papers in journals and conferences in these areas. He is also a Committee Member of the China Computer Federation.

ZHONGNAN WEI received the B.S. degree in automation from the Guangdong University of Petrochemical Technology, China, in 2018. He is currently pursuing the master's degree with the School of Computer and Information Engineering, Xiamen University of Technology, China. His research interests include machine learning, rough set and granular computing, cloud computing, and information security.

YUMIN CHEN (Member, IEEE) received the M.E. degree in computer application technology from Nanchang University, China, in 2005, and the Ph.D. degree in pattern recognition and intelligence system from Tongji University, China, in 2010. From November 2018 to November 2019, he was a Visiting Scholar, supported by the Education Department of Fujian Province, with the Department of Mathematics and Statistics, Mississippi State University, USA. He is currently a Professor with the College of Computer and Information Engineering, Xiamen University of Technology. He has published over 50 papers in refereed journals and conferences. His research interests include machine learning, rough sets, and granular computing.



CHAO TANG was born in Hefei, Anhui, China. He received the M.S. degree from Shanxi University, Taiyuan, China, in 2009, and the Ph.D. degree in artificial intelligence from Xiamen University, Xiamen, China, in 2014. He is currently an Associate Professor with the School of Artificial Intelligence and Big Data, Hefei University, China. His research interests include machine learning, computer vision, and human action recognition.

YUPING SONG received the Ph.D. degree from Peking University, China, in 2007. She is currently an Associate Professor with the School of Mathematical Sciences, Xiamen University, Xiamen, China. Her research interests include differential geometry, fuzzy modeling, and granular computing.