Journal of Network Intelligence ©2024 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **9**, Number **4**, November **2024**

Identification and Prediction of Psychological Stress in Athletes Based on Fuzzy Clustering

Yong-Xing Wang*

Sports Teaching Department
Shangqiu Institute of Technology, Shangqiu 476000, P. R. China
Yongxingshangqiu@163.com

Fang Li

College of Sciences
Trinity University of Asia, 275 E. Rodriguez Sr. Blvd., Manila, Philippines
kn8636@163.com

*Corresponding author: Yong-Xing Wang
Received December 31, 2023, revised March 13, 2024, accepted June 18, 2024.

Abstract. *In the modern environment where the pressure of work and life is increasing day by day, the problems brought by psychological stress to the physical and mental health of athletes are becoming more and more significant, therefore, the accurate identification and prediction of psychological stress is particularly important. Aiming at the problem that current psychological stress recognition methods cannot accurately predict stress, this paper proposes a fuzzy clustering-based method for recognizing and predicting psychological stress in athletes. The method firstly improves the traditional Fuzzy C-means clustering (FCM) based on the target C-mean, based on the concept of information granularity, and combines the interclass coupling and separation indexes of the clustered samples to construct the validity function for discriminating the classification effect, which effectively enhances the efficiency of the algorithm for solving the optimal value in the iterative process. Then, according to the structural characteristics of the athlete's ECG pressure monitoring equipment, extract the different transmission characteristics of its signal electromagnetic wave, use the optimized FCM algorithm to cluster the irregularly distributed ECG feature data samples, constitute different levels of feature vectors, so as to construct the variable prediction model and realize the rapid identification of the athlete's psychological pressure. The experimental results show that the IPPSA method can accurately recognize the category and number of data samples. As for 500 test samples, the number of misclassification of IPPSA method is only 19, and the classification error rate is 3.8%, while the RMSE is 0.0291. When recognizing the stress state, the AUC is 0.78, and the accuracy rate is 94.01%.*
Keywords: psychological stress; fuzzy clustering; information granularity; feature vector; validity function

1. **Introduction.** In recent years, stress has become a widespread societal problem affecting people of all ages, genders or industries in all countries around the globe [1, 2]. Athletes, in particular, are more prone to various emotions due to a sharp rise in stress caused by, among other things, increased uncertainty in their lives. However, when the emotional threshold is higher, the likelihood of developing stress increases, creating a vicious cycle [3, 4]. According to the United Nations Secretary General's World Health Day statement, approximately 1 billion people worldwide are suffering from mental health

problems, with depression and anxiety disorders being among the most common psychiatric disorders [5]. The results of a recent survey collected by the Depression Research Institute in collaboration with the People's Daily client among patients with depression; the questionnaire showed that 86% of the patients believed that emotional stress was the main cause of depression or anxiety. Psychological stress may bring about a variety of physical and psychological adverse reactions [6, 7, 8], such as anxiety, depression, insomnia, fatigue, indigestion, headache, etc. It may also lead to serious health problems, such as hypertension, cardiovascular diseases, immune system disorders, etc., if they are in a state of high psychological stress for a long time. Thus, how to recognize and predict the psychological stress of athletes is of great research significance.

1.1. **Related Work.** Due to the adverse effects of stress on people's physical and mental health, more and more researchers from all over the world have begun to pay attention to the field of psychological stress recognition and prediction. Healey and Picard [9] proposed a method to collect and analyze drivers' physiological data to determine the relative stress level of the human body in real driving scenarios, and the results showed that the electrical skin signals and the heart rate are the most closely correlated. This study demonstrated for the first time the feasibility of using physiological signals to detect psychological stress in humans. Setz et al. [10] in their study induced stress in subjects by asking them to complete a social stress test and a mathematical task within a restricted time frame, and achieved a classification accuracy of 82.8% for stress under the SVM model. Misra et al. [11] used the Stroop test as a stressor and collected data such as pulse wave, skin electricity, and skin temperature. Olusoga et al. [12] proposed a system for recognizing emotional stress and achieved up to 82.7% accuracy in classifying stress. Liu and Jiang [13] first converted the EEG data segments of athletes into feature vectors obeying polynomial distributions, and then fused them into response-level feature vectors. Cormier and Zizzi [14] used the generalized estimating equations in the generalized linear model for data classification experiments and achieved up to 74.5% accuracy.

Subsequently, Goyal et al. [15] proposed an algorithm based on multilayer perceptron, generalized regression neural network and adaptive neuro-fuzzy system. Ji et al. [16] proposed a hybrid genetic algorithm combining an artificial neural network with a support vector machine for the recognition of stress. Flood and Keegan [17] recorded the EEG data of the subjects and classified the data by SVM. Nixdorf et al. [18] achieved good accuracy in predicting the stress level of the user through Bayesian network. Ruiz et al. [19] extracted 63 features from physiological signals and used them as inputs to the Random Forest algorithm and finally achieved 72% accuracy. Schaal et al. [20] investigated the detection of stress and emotion based on physiological signals and implemented binary classification using various machine learning algorithms. multiple machine learning algorithms to achieve a binary classification task (stressed and unstressed states). Hofmockel et al. [21] combined sensor signals from different vehicles to enhance the accuracy of road condition prediction. Choi and Lee [22] fused the eigenvectors of multiple modalities through polynomial sampling to ensure compatibility with other learning models. Johnson and Ivarsson [23] fused features from multiple sensor signals in conjunction with an Embrace Net network, which was used to improve the accuracy of target recognition. Xie et al. [24] proposed a multimodal emotion recognition method that combines a polynomial sampling process to fuse audio, video, and text modalities, which outperforms unimodal recognition models. On the ground of this, Heredia et al. [25] offered an adaptive emotion recognition architecture that can cope well with the problem of missing data. Li and Liu [26] suggested a one-dimensional convolutional neural network for stress recognition, which extracts features from the raw data of different modalities (ECG signals, dermatoelectric signals, etc.). Xu et al. [27] used Fuzzy C-mean clustering (FCM) mean clustering method to cluster the samples to get the clustering center, and then used ordered weighting operator to cluster the center to get a new method to make the prediction of athlete's psychological stress, but the prediction effect is not good. Li et al. [28] proposed a model for recognizing psychological stress in depressed patients based on machine learning and fuzzy K-means clustering, but it could not be clustered effectively, resulting in inefficient recognition.

1.2. **Motivation and contribution.** However, the existing research on psychological stress recognition ignores the characteristics of psychological stress itself, resulting in poor generalization ability and low accuracy.

To address the above problems, this paper first optimizes the target C-mean fuzzy clustering algorithm, introduces the K-means clustering for the initialization of the category center, so that the algorithm can obtain the clustering center that is consistent with the distribution of the sample space, and the sample clustering provides an explicit indicator of the compliance rules, and effectively enhances the algorithm's efficiency of the optimal value solution in the iterative process. Then the different transmission characteristics of electromagnetic waves of cardiac pressure signals of athletes are extracted, and the optimized FCM algorithm is used to find the objective function, so as to obtain the affiliation matrix and the clustering center vector. The extracted data are all in the same storage space, and the dispersion degree of the feature data is measured using the improved fuzzy clustering algorithm before the intelligent recognition, so as to carry out effective recognition. The experimental results show that the IPPSA method designed in this paper has a high accuracy rate and can quickly recognize the psychological stress of athletes.

2. Basic theoretical analysis.

2.1. **Fuzzy c-means clustering.** The FCM algorithm is currently the most widely used fuzzy clustering algorithm, probably for the following reasons: first, it divides the clustering into nonlinear problems with constraints, and the fuzzy division and clustering problems are obtained by optimizing the solution [28], which makes the computation simplified and has a wide range of applications. The clustering process is indicated in Figure 1. Compared with other clustering algorithms, FCM clustering can resist the interference of outliers or noise data because it blurs the distance between each point and each clustering center.

3. Basic theoretical analysis.

3.1. **Fuzzy c-means clustering.** The given set of n samples $Y = \{y_1, y_2, \ldots, y_m\}$ is divided into c classes, each sample belongs to c different domains with a certain degree of affiliation, and the clustering centers of the c classes are called G, the affiliation matrix is called F, and η_{ij} denotes the degree of affiliation of the j-th sample to the i-th class.

The objective function for fuzzy c-mean clustering is as follows.

$$J_m(F, G) = \sum_{j=1}^{m} \sum_{i=1}^{c} (\eta_{ij})^n (d_{ij})^2 \tag{1}$$

where n is the weighting index; d_{ij} is the distance between sample y_j and the clustering center g_i of class i, which is called the Euclidean distance. The smaller the value of $J_m(F, G)$, the better the clustering effect.

With this objective function the two-by-two differences between the data points and the clustering centers can be summed, so minimizing the objective function can find the most compact clusters. Under the constraints, the minimum solution of the objective function is solved according to the Lagrange multiplier method, which is defined as Equation (2).

$$\min J_m(F,G) = \sum_{j=1}^{m}\sum_{i=1}^{c}(\eta_{ij})^n (d_{ij})^2 - \xi\left(\sum_{i=1}^{c}(\eta_{ij})^n - 1\right) \tag{2}$$

where ξ is the Lagrange multiplier.

The clustering criterion of the FCM, i.e., determining F, G, allows convergence in terms of minimized iterations. At this point we use the Lagrange multiplier method to find η_{ij}.

$$g_i = \frac{\sum_{j=1}^{m}(\eta_{ij})^n y_j}{\sum_{j=1}^{m}(\eta_{ij})^n} \tag{3}$$

$$\eta_{ij} = \sum_{l=1}^{c}\left(\frac{d_{ij}}{d_{lj}}\right)^{\frac{2}{n-1}} \tag{4}$$

From the above equation, the FCM algorithm needs to be optimized by constantly modifying and optimizing the clustering center matrix and the affiliation matrix. The advantage of the FCM algorithm is that the computational process is simple, fast running, intuitive, and can be processed and analyzed for large data volume sample sets. ECG signals are composed of ECG waveforms, and FCM clustering can identify different ECG waveforms and divide them into different clusters.

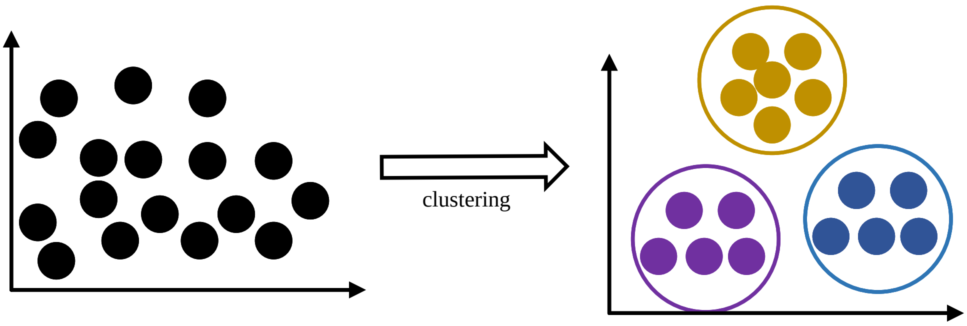

Figure 1. The clustering process

3.2. **Support vector machine.** Support vector machine [29] is a pattern recognition algorithm based on the basis of statistical learning theory, and its goal is to find a hyperplane that can separate the samples and has the maximum spacing, as shown in Figure 2, which is the desired separation hyperplane.

Given a training dataset $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$ in feature space, where $x_j \in \mathbb{R}^m$, $y_j \in \{+1, -1\}$, $j = 1, 2, \ldots, M$, y_j is the j-th feature vector and y_j is the category labeling, when $y_j = +1$, it denotes a positive case and when $y_j = -1$, it denotes a negative case. For a given training dataset S and separating hyperplane $h \cdot y + a = 0$, first define the geometric interval of the hyperplane about a sample point (x_j, y_j) as follows.

$$\alpha_j = y_j\left(\frac{h}{\|h\|}\cdot y_j + \frac{a}{\|h\|}\right) \tag{5}$$

The minimum value of the geometric interval of the partition hyperplane for all sample points is $\alpha = \min_{j=1,2,\ldots,M}\alpha_j$. α denotes actually the distance between the support vector and the hyperplane in SVM. Under the above definition, the problem of solving the SVM model for a partitioned hyperplane with maximum interval can be defined as an optimization problem with constraints, as indicated in Equation (6).

$$\max_{h,a} \alpha \quad \text{s.t.} \quad y_j\left(\frac{h}{\|h\|} \cdot x_j + \frac{a}{\|h\|}\right) \geq \alpha, \quad j = 1, 2, \ldots, M \tag{6}$$

Simplification yields: $y_j(h \cdot x_j + a) \geq 1, j = 1, 2, \ldots, M$.

For a linearly indivisible problem, it is usually necessary to introduce relaxation variables ε_j and Lagrange multipliers to transform Equation (6) above into a dyadic problem, and then solve for the optimal separating hyperplane. In addition, a penalty factor C needs to be introduced, and the optimization problem above is transformed into Equation (7).

$$\min_{h,a,\varepsilon} \frac{1}{2}\|h\|^2 + C\sum_{j=1}^{n} \varepsilon \quad \text{s.t.} \quad y_i(h \cdot x_j + a) \geq 1 - \varepsilon_j, \quad \varepsilon_j \geq 0, \quad j = 1, 2, \ldots, M \tag{7}$$

When C is smaller, the model complexity is lower and is prone to underfitting. With the Lagrange operator, the above equation can be transformed into Equation (8).

$$\min \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \beta_i\beta_j y_i y_j (x_i \cdot x_j) - \sum_{j=1}^{M} \beta_j \quad \text{s.t.} \quad \sum_{j=1}^{M} \beta_j y_j = 0, \quad 0 \leq \beta_j \leq C, \quad j = 1, 2, \ldots, M \tag{8}$$

With the Sequential Minimum Optimization (SMO) algorithm, the optimal solution can be computed as follows.

$$\beta^* = (\beta_1^*, \beta_2^*, \ldots, \beta_M^*)^T \tag{9}$$

$$h^* = \sum_{j=1}^{M} \beta_j^* y_j x_j \tag{10}$$

The final categorical hyperplane decision function obtained is as follows.

$$f(x) = \text{sign}(h^* \cdot x + a - \sum_{i=1}^{M} \beta_i^* y_i (x_i \cdot x_j)) \tag{11}$$

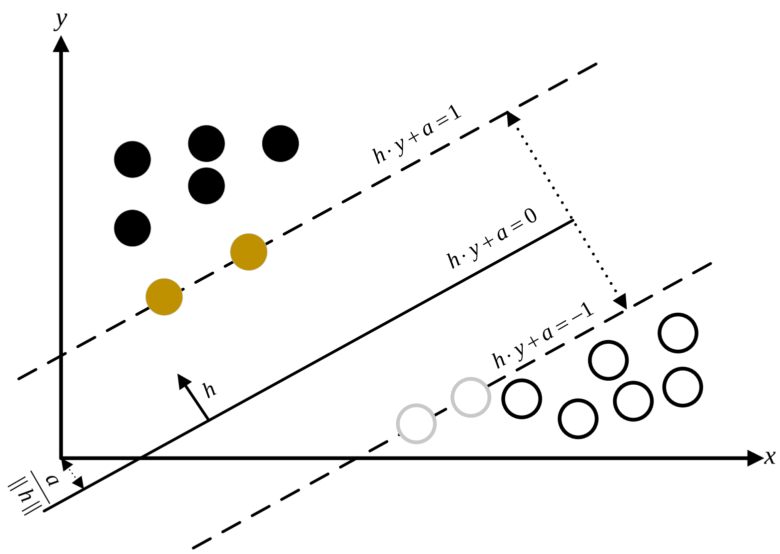

Figure 2. Support vector machine

3.3. **Improved fuzzy clustering algorithm.** Although the traditional FCM algorithm is widely used, it is more sensitive to the initial value when analyzing the psychological data of athletes, which leads to slower convergence, and it is easy to fall into the local optimum in the iterative process, which affects the accuracy of data analysis and prediction. Therefore, this paper introduces the affiliation function, which adds a non-qualitative description of the sample category and establishes a more suitable mapping relationship between the object and the objective world. In addition, the method can automatically extract the features from the athletes' psychological data, and then realize the autonomous classification of the samples. Relied on the discussion of the commonly used fuzzy cluster analysis algorithm based on objective function, the algorithm is improved with the application scenario of analyzing the information of athletes' psychological data. The flow of the enhanced algorithm is implied in Figure 3.

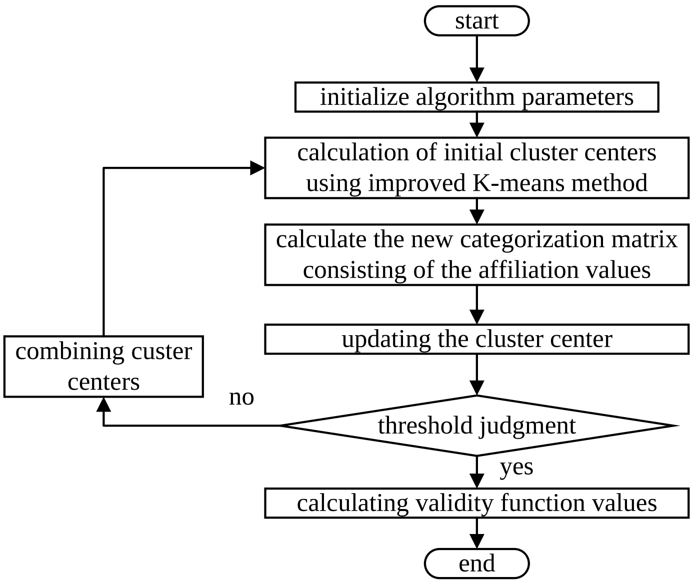

Figure 3. Improved fuzzy clustering algorithm process

3.4. **Clustering Center Calculation and Effectiveness Function in FCM.** The initial clustering center is first calculated using the K-means method with the expression:

$$E = \frac{1}{j} \sum_{j=p\in C_j} |p - n_j|^2 \tag{12}$$

The sum of the squared differences of all the objects in the two-dimensional space to the center of the sample is used as the error discriminant E of K-means, while p is the mapping of the input sample in the two-dimensional space, and n_j is the cluster center of C_j. The method is used to initialize the cluster center of FCM. The introduction of this method can effectively enhance the initialization of cluster centers in FCM.

In addition, to measure the clustering impact of the FCM algorithm, an effectiveness function on the ground of information granularity is introduced. The information granularity can characterize the coupling of samples among classes, and it mainly contains the concepts of coupling degree $Cd(c)$ and separation degree $Td(c)$.

$$Cd(c) = \frac{1}{m} \sum_{i=1}^{c} \sum_{j=1}^{m} f_{ij}^{n} d_{ij}^{2} \tag{13}$$

where $i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, m$, f_{ij} is the degree of affiliation and d_{ij} is the Euclidean distance.

$$Td(c) = \frac{\sum_{i=1,l=1,i\neq l}^{c} d_{il}^2}{[c(c-1)/2]} \tag{14}$$

where $i, l = 1, 2, \ldots, c$.

Based on Equation (13) and Equation (14), the effectiveness function to measure the clustering effect can be obtained as Equation (15).

$$VD(c) = \beta Cd(c) + (1-\beta)\frac{1}{Td(c)} \tag{15}$$

where β is the weight adjustment factor between coupling and dispersion.

According to Figure 3, the improved FCM will discriminate the classification effect according to the validity function VD, so that the distance between samples within a class is as small as possible, while the distance between cluster centers between categories is as large as possible. The distance between cluster centers between categories is determined as follows.

$$D(i,j) = g_i - g_j, i \neq j \tag{16}$$

In addition, to avoid the influence of data noise on the discrimination of fuzzy matrix affiliation, the objective function is also improved in the article.

$$\min J_{FCM} = \sum_{l=1}^{m}\sum_{i=1}^{c} f_{il}^n (d_{il})^2 + \sum_{i=1}^{c}\mu_i \sum_{l=1}^{m}(1-f_{il})^n \tag{17}$$

where μ_i is the relaxation factor, which reduces the constraint of the original loss function on the degree of affiliation. The parameter is tabulated as follows.

$$\mu_i = L\frac{\sum_{i=1}^{m} f_{il}d_{il}^2}{\sum_{l=1}^{m} f_{il}^n} \tag{18}$$

where L is a constant. Then the improved FCM algorithm parameter update method is as follows.

$$f_{il}^{(s+1)} = \left(\sum_{j=1}^{c} d_{jl}^{(s)}\right)^{\frac{2}{n-1}} \tag{19}$$

$$g_i^{(s+1)} = \frac{\sum_{l=1}^{m}\left(f_{il}^{(s)}\right)^n \cdot x_l}{\sum_{l=1}^{m}\left(f_{il}^{(s)}\right)^n} \tag{20}$$

where f_{ij} is the degree of affiliation and g_i is the clustering center.

4. **Identification and prediction of psychological stress in athletes based on improved fuzzy clustering.**

4.1. **Athletes' ECG signal characterization data extraction.**

The traditional method of recognizing athletes' psychological stress still has certain defects, so this paper integrates the FCM algorithm and proposes a brand-new method of recognizing athletes' psychological stress. The method first extracts the athlete's ECG signal feature data, and then SVM algorithm processes the extracted feature data to form feature vectors at different levels. The feature vectors are used to study linear or planar data, and the dispersion of the data is verified through the metric clustering covariance matrix, so as to realize the accurate recognition and prediction of athletes' psychological stress. The overall process of identification is indicated in Figure 4.

According to the structural characteristics of the athlete's cardiac stress monitoring equipment, different transmission characteristics of its signal electromagnetic wave are extracted. In a single insulating medium, electromagnetic wave energy flows along the electromagnetic wave transmission direction, at which time the flow speed is the electromagnetic wave transmission wave speed. Assuming the existence of a parallel plane x and y, the transmission speed along the z-axis is as follows.

$$v = \frac{dz}{dt} = \frac{1}{\sqrt{\lambda\vartheta}} \tag{21}$$

where v denotes the output velocity, dz/dt denotes the ECG data analyzed, λ denotes the magnetic permeability of the propagation medium, and ϑ denotes the permittivity of the propagation medium.

And the transmission theorem refers to the electromagnetic wave transmitted in the positive direction of the z-axis as the incident wave, then any location node in the insulated space, its electromagnetic field energy density is Equation (22).

$$\rho = \lambda W E_y^2 = \vartheta g^2 \tag{22}$$

where W and E denote fixed parameters in Maxwell's fundamental equations, y denotes electromagnetic wave transmission. The energy flow density vector in an insulating medium can be described by $V = WE = g\rho$.

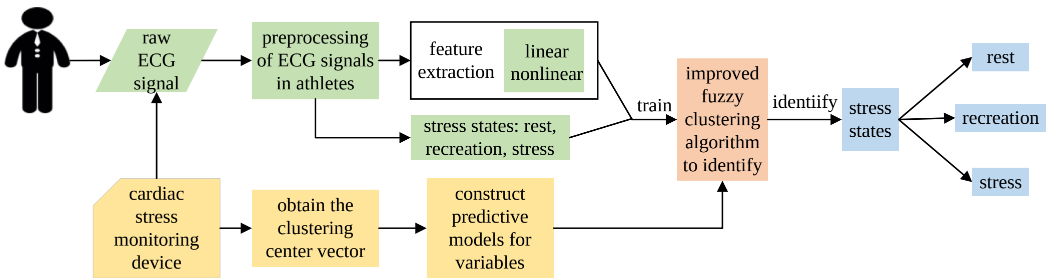

Figure 4. Fuzzy clustering based psychological stress identification process for athletes

There is an attenuation phenomenon of electromagnetic wave propagation in a conductor, so the transmission characteristics of electromagnetic wave in a uniform conducting medium are described as follows.

$$V^2 E - i\rho t\lambda E + \rho^2 \vartheta \lambda E = 0 \tag{23}$$

where i denotes the path and t denotes the conductivity. It is known that there are $s/\rho\vartheta \geq 1, \tan\phi \geq 1$, in the good conductor, and transmission parameters are calculated as Equation (24).

$$\beta \approx \sqrt{\frac{\rho\lambda t}{2}}, \alpha \approx \sqrt{\frac{\rho\lambda t}{2}}, g \approx \sqrt{\frac{2\rho}{\lambda t}}, z \approx \sqrt{\frac{\rho\lambda}{t}} e^{\frac{i\pi}{4}} \tag{24}$$

where β denotes the electromagnetic wave attenuation coefficient, α denotes the electromagnetic wave propagation phase coefficient, g denotes the electromagnetic wave propagation velocity, and z denotes the electromagnetic wave propagation wave impedance.

The result of the calculation, w indicates the depth of penetration, reflecting the relationship between the parameters α and β, and realizes the extraction of the characteristics of partial discharge ultra-high frequency ECG electromagnetic wave transmission and ECG signals.

4.2. **Clustering psychological feature data.** Based on the extracted features of the athlete's ECG pressure signals, the extracted feature data are first processed using SVM algorithm to form feature vectors at different levels. Then the feature vectors and the improved fuzzy clustering algorithm is used to find the objective function, so as to obtain the affiliation matrix and the clustering center vector. The extracted data are all in the same storage space, and the improved fuzzy clustering algorithm is used to measure the degree of dispersion of the feature data before intelligent recognition, so as to carry out effective classification. Assuming that the ECG device signal eigenvector, denoted by $S(S_1, S_2, ..., S_P)$, the clustering center vector of the feature data is $P = [p_1, p_2, ..., p_m]^T$, where m denotes the number of clusters, the affiliation matrix of the ECG stress feature sample data is $H = [h_{fg}]_{m\times n}$, n denotes the number of samples, f and g denote the location of the feature data points. In practical clustering, the affiliation matrix satisfies $h_{fg} \in [0, 1]$, $\sum_{f=1}^{m} h_{fg} = 1$, $1 < g < n$.

The property metrics of fuzzy clustering in ECG stress signal recognition are as follows.

$$U(X, P, H) = \sum_{f=1}^{m}\sum_{g=1}^{n}(h_{fg})^j L_{fg}^2 \tag{25}$$

where i denotes the fuzzy index parameter describing the degree of fuzzy clustering, $X = (x_1, x_2, \cdots, x_n)$ is a set of psychological stress characteristic data series, L_{fg}^2 denotes the squared inner product paradigm, which is the distance from any data point x_g to the clustering center p_f. The computation equation of L_{fg}^2 is as follows.

$$L_{fg}^2 = x_g - p_f\mathbf{A}_f^2(x_g - p_f)^\top\mathbf{A}_f(x_g - p_f) \tag{26}$$

where A_f denotes the positive definite symmetric matrix, which is mainly determined by the clustering covariance matrix.

The Lagrange multiplication optimization Equation (26) is adopted to satisfy the minimum (H, P) condition as follows.

$$h_{fg} = \frac{1}{\sum_{i=1}^{m}\left(\frac{L_{fg}}{L_{ig}}\right)^{i-1}} \tag{27}$$

$$p_f = \frac{\sum_{g=1}^{n}(h_{fg})^i x_g}{\sum_{g=1}^{n}(h_{fg})^i} \tag{28}$$

In the actual processing, after each parameter meets the conditions of Equation (28), set the number of iterations as $d = 1, 2, 3, \ldots, D$ to update the clustering center p_f, and calculate the fuzzy covariance matrix E_f.

$$E_f = \frac{\sum_{g=1}^{n}(h_{fg})^i(x_g - p_f)(x_g - p_f)^\top}{\sum_{g=1}^{n}(h_{fg})^i} \tag{29}$$

Updating the fuzzy division matrix H, if the calculation meets $\|H^d - H^{d-1}\| < \beta$, then terminate the operation. β denotes any positive number set by the person. If the above conditions are not met, then set the number of iterations to $d = d + 1$, re-update the clustering center, repeat the calculation process until the conditions are met, and output the results, and use the output results as the input of the athlete's psychological stress recognition.

4.3. **Intelligent identification and prediction of psychological stress in athletes.** Assuming that a set of cardiac stress signal feature vectors extracted using by SVM algorithm is denoted by $S(S_1, S_2, \cdots, S_P)$, there may be a corresponding linear or nonlinear relationship between these data, so there is the Equation (30).

$$S_1 = F_1(n, m)(S_2) \tag{30}$$

where F denotes the closeness when the domain is discrete. There may also be a one-to-many linear relationship, or a nonlinear relationship, in which case the following equation exists.

$$S_1 = F(S_2, S_3, \ldots, S_P) \tag{31}$$

Different mathematical models are established to distinguish different interrelationships in order to enhance the effect of pressure recognition. Assuming a prediction parameter is S_j, build a variable prediction model as shown in the following equation.

$$S_j = F(S_i, t_0, t_i, t_{ij}, t_{il}) + e \tag{32}$$

where t_0, t_i, t_{ij}, and t_{il} represent the prediction model parameters obtained through training sample data, S_j represents the predicted variable, S_i represents the predictor variable, and e represents the prediction error.

Using the prediction model for data training and testing of two processes to obtain the athlete psychological pressure data center, so as to complete the improved fuzzy clustering algorithm based on FCM to achieve reliable identification and prediction of the athlete's psychological pressure.

5. Experiment and analysis.

5.1. **Analysis of clustering effect.** For the purpose of estimating the performance of the fuzzy clustering-based method for recognizing and predicting athletes' stress, this paper conducts a comparative experiment on the WESAD dataset with athletes' psychological stress data. WESAD [30] is an open-source psychological stress dataset, which contains 16 psychological states, and the parameters of the dataset selected for this paper are as follows. The total number of samples has 500; there are three kinds of psychological stress states of athletes: resting state, recreational state and stress state, and the number of samples in each category is 150; and there are 8 stress characteristic parameters. The comparison models for the experiments were trained in Python v3.7.2 environment. For ease of description, the literature [31] is denoted as SDBFI, the literature [?] is denoted as AILKA, and the algorithm in this paper is denoted as IPPSA.

According to the actual application scenario of the algorithm, before clustering and analyzing the psychological stress data, the number of categories c is not a definite value, so it also needs to be determined according to the algorithm's validity function value. Table 1 gives the validity function values of the algorithm when the dataset is divided into different categories in the simulation process, and only 7 categories are listed due to space limitations. It can be seen that when c = 3, the algorithm can get the optimal effectiveness function value, about 0.4318, and the number of categories is also consistent with the actual number of categories of the data set.

The dataset in Table 1 is normalized and mapped to a two-dimensional space, and then the pressure data are clustered using the algorithm in the paper, and the results obtained

Table 1. Validity function values corresponding to different categories of numbers.

Number of categories	Validity function value
1	0.8692
2	0.7931
3	0.4318
4	0.5963
5	0.6891
6	0.8712
7	0.9154

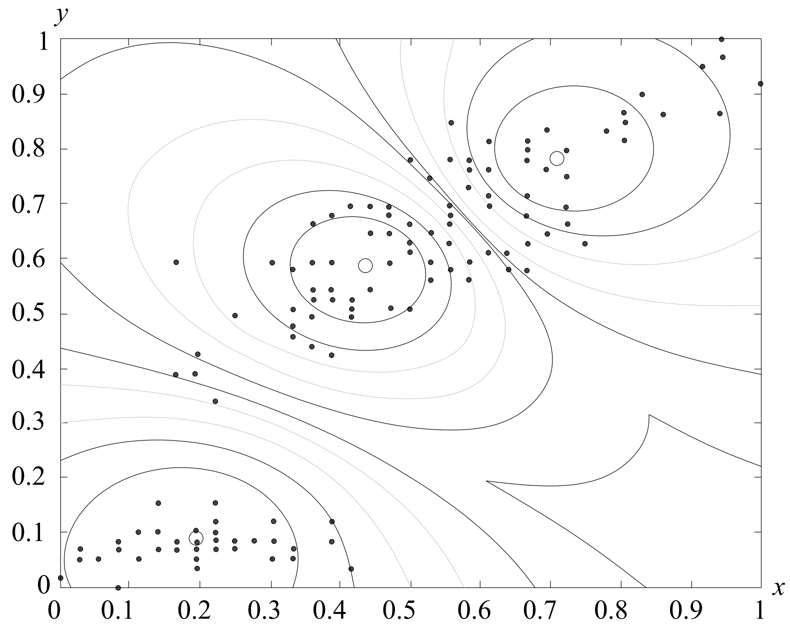

Figure 5. Clustering effect

are shown in Figure 5. The hollow circles are the actual clustering centers and the arcs are the boundaries of the categories. It can be seen that the arcs divide all the samples into three categories and there is no overlapping between the clusters. This proves that the proposed algorithm can categorize all the data in the dataset clearly.

Table 2. Algorithm clustering performance comparison

Norm	Traditional FCM	Improvement of FCM
Statistic of correct identification / times	41	48
Correct rate (%)	82	96
Average run time / s	297	204

For the clustering algorithm, the first step is to classify a bunch of unorganized data into correct classes. Table 2 shows the statistical results of the algorithm that can correctly classify the experimental data into three categories in 50 runs. As can be seen from the table, the correct rate of the algorithm is 98%, which is 14% higher than that of the traditional algorithm; and the average running time is reduced to 204 s, which is 31.3% shorter than that of the traditional algorithm.

Table 3 counts the classification accuracy of all samples when the amount of categories $c = 3$. For 500 test samples, the amount of misclassifications of IPPSA is 19, the classification error rate is 3.8%, and the root mean square error (RMSE) value is 0.0291. Compared with traditional FCM algorithm, error rate decreases by 8%, and RMSE value decreases by 0.1403.

Table 3. Algorithm classification accuracy performance comparison

Norm	**Traditional FCM**	**Improvement of FCM**
Number of misclassifications per	59	19
Error rate (%)	11.8	3.8
RMSE	0.1694	0.0291

5.2. **Comparative test analysis.** For the purpose of estimating the performance of the IPPSA suggested in this article, this article compares and analyzes with the SDBFI and AILKA models, and identifies the resting state, recreational state, and pressure state in the dataset, and the average accuracy of 100 validation tests of the experimental model for each type of Receiver Operating Characteristic (ROC) Curve is adopted as one of the evaluation indexes of the model performance. Due to space limitation, only the ROC curves for the stress state are shown here. The ROC curves and Area under Curve (AUC) values of IPPSA and the comparison model for recognizing different psychological stress states are shown in Fig. 6, in which the diagonal dashed line indicates the result of random prediction, whose AUC is 0.5. When recognizing the stress state, AUCs of SDBFI, AILKA, and IPPSA are 0.62, 0.75, and 0.78, respectively, and the IPPSA is the closest to the random prediction result. The SDBFI and AILKA models had poor overall identification and prediction results.

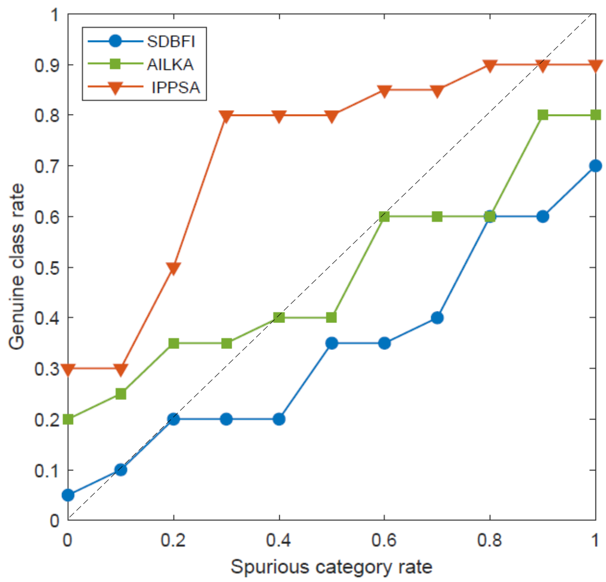

Figure 6. ROC curves and AUC values for different models

Figure 7 compares the recognition and prediction accuracy of SDBFI, AILKA and IPPSA for the three states. The accuracy of the experimental comparison models in recognizing the three states is consistent with the ROC curve and AUC performance.

When in the recreational state, the recognition accuracy of the three models was low. Only when in the stressful state, SDBFI, AILKA and IPPSA recognition achieves better results. Meanwhile, comparing the recognition accuracy of the three models, it is found that SDBFI has a lower accuracy of 80.24% compared with AILKA and IPPSA models in recognizing the stressful state, while IPPSA model has the highest accuracy of 94.01%.

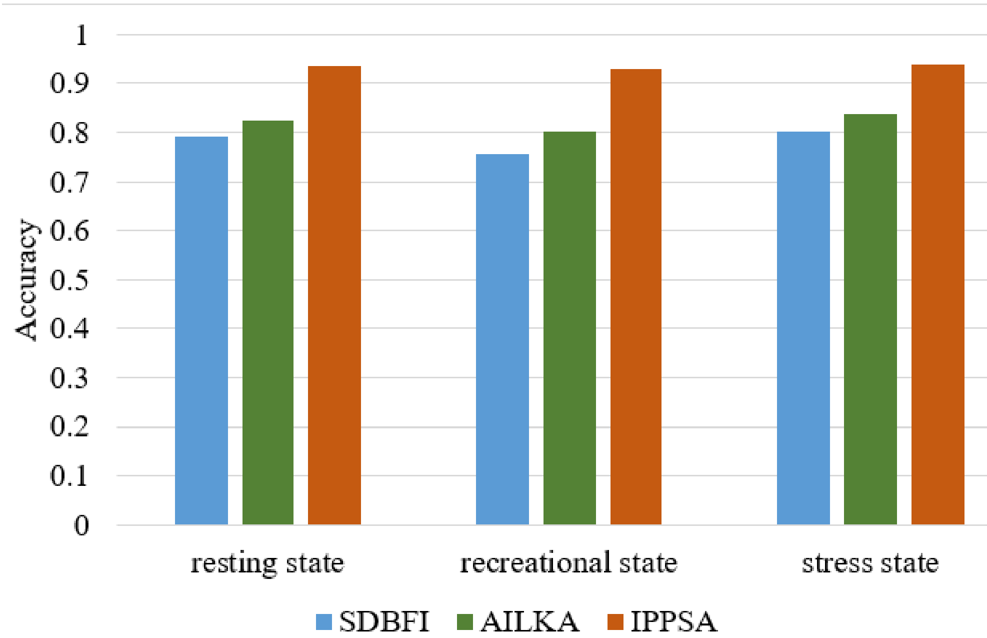

Figure 7. Comparison of recognition and prediction accuracy of different models

6. **Conclusion.** In the modern environment where the pressure of work and life is increasing day by day, the problems brought by psychological stress to the physical and mental health of athletes are becoming more and more serious, and the research on the identification of psychological stress has been widely concerned by researchers at home and abroad. In this context, to address the problem of low recognition efficiency of the current psychological stress recognition methods, this paper first introduces K-means clustering for the initialization of category centers, so that the algorithm can obtain the cluster centers that are consistent with the distribution of the sample space, improve the traditional FCM algorithm, and then improve the algorithm's efficiency of solving the optimal value in the iterative process. Then, according to the characteristics of the athlete's ECG pressure monitoring equipment, different transmission characteristics of its signal electromagnetic wave are extracted, the optimized FCM algorithm is used to obtain the objective function, so as to obtain the affiliation matrix and the clustering center vector, and the improved fuzzy clustering algorithm is adopted to measure the dispersion degree of the feature data and construct the variable prediction model, so as to realize the rapid identification of the athlete's psychological pressure. The experimental outcome indicates that the IPPSA method offered in this article has a high AUC value and accuracy rate, and can greatly realize the accurate prediction of athletes' psychological stress.

REFERENCES

[1] R. Maddison, and H. Prapavessis, "A psychological approach to the prediction and prevention of athletic injury," *Journal of Sport and Exercise Psychology*, vol. 27, no. 3, pp. 289-310, 2005.

[2] J. Peacock, and W. Whang, "Psychological distress and arrhythmia: risk prediction and potential modifiers," *Progress in Cardiovascular Diseases*, vol. 55, no. 6, pp. 582-589, 2013.

[3] D. Roh, J. R. Chao, D. H. Kim, K. H. Yoon, J. H. Jung, C. H. Lee, J.-H. Shin, M. J. Kim, C. H. Park, and J. H. Lee, "Psychological stress as a measure for treatment response prediction in idiopathic sudden hearing loss," *Journal of Psychosomatic Research*, vol. 102, pp. 41-46, 2017.

[4] A. Moatamedy, A. Borjali, and M. Sadeqpur, "Prediction of psychological well-being of the elderly based on the power of stress management and social support," *Iranian Journal of Ageing*, vol. 13, no. 1, pp. 98-109, 2018.

[5] L. L. Craft, and D. M. Landers, "The effect of exercise on clinical depression and depression resulting from mental illness: A meta-analysis," *Journal of Sport and Exercise Psychology*, vol. 20, no. 4, pp. 339-357, 1998.

[6] N. Sharma, and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 3, pp. 1287-1301, 2012.

[7] H. Gustafsson, S. S. Sagar, and A. Stenling, "Fear of failure, psychological stress, and burnout among adolescent athletes competing in high level sport," *Scandinavian Journal of Medicine & Science in Sports*, vol. 27, no. 12, pp. 2091-2102, 2017.

[8] M. G. Hutchison, L. Mainwaring, A. Senthinathan, N. Churchill, S. Thomas, and D. Richards, "Psychological and physiological markers of stress in concussed athletes across recovery milestones," *Journal of Head Trauma Rehabilitation*, vol. 32, no. 3, pp. E38-E48, 2017.

[9] J. A. Healey, and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156-166, 2005.

[10] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 410-417, 2009.

[11] S. Misra, C. Yoon, K. J. Kim, R. Managuli, R. G. Barr, J. Baek, and C. Kim, "Deep learning-based multimodal fusion network for segmentation and classification of breast cancers using B-mode and elastography ultrasound images," *Bioengineering & Translational Medicine*, pp. e10480, 2022.

[12] P. Olusoga, J. Butt, K. Hays, and I. Maynard, "Stress in elite sports coaching: Identifying stressors," *Journal of Applied Sport Psychology*, vol. 21, no. 4, pp. 442-459, 2009.

[13] Y. Liu, and C. Jiang, "Recognition of shooter's emotions under stress based on affective computing," *IEEE Access*, vol. 7, pp. 62338-62343, 2019.

[14] M. L. Cormier, and S. J. Zizzi, "Athletic trainers' skills in identifying and managing athletes experiencing psychological distress," *Journal of Athletic Training*, vol. 50, no. 12, pp. 1267-1276, 2015.

[15] A. Goyal, S. Singh, D. Vir, and D. Pershad, "Automation of stress recognition using subjective or objective measures," *Psychological Studies*, vol. 61, pp. 348-364, 2016.

[16] J. Ji, T. Wan, D. Chen, H. Wang, M. Zheng, and Z. Qin, "A deep learning method for automatic evaluation of diagnostic information from multi-stained histopathological images," *Knowledge-Based Systems*, vol. 256, pp. 109820, 2022.

[17] A. Flood, and R. J. Keegan, "Cognitive resilience to psychological stress in military personnel," *Frontiers in Psychology*, vol. 13, pp. 809003, 2022.

[18] I. Nixdorf, R. Frank, and J. Beckmann, "An explorative study on major stressors and its connection to depression and chronic stress among German elite athletes," *Advances in Physical Education*, vol. 5, no. 04, pp. 255, 2015.

[19] M. C. Ruiz, C. Robazza, A. Tolvanen, S. Haapanen, and J. L. Duda, "Coach-created motivational climate and athletes' adaptation to psychological stress: Temporal motivation-emotion interplay," *Frontiers in Psychology*, vol. 10, pp. 617, 2019.

[20] K. Schaal, M. Tafflet, H. Nassif, V. Thibault, C. Pichard, M. Alcotte, T. Guillet, N. El Helou, G. Berthelot, and S. Simon, "Psychological balance in high level athletes: gender-based differences and sport-specific patterns," *PloS One*, vol. 6, no. 5, pp. e19007, 2011.

[21] J. Hofmockel, J. Masino, J. Thumm, E. Sax, and F. Gauterin, "Multiple vehicle fusion for a robust road condition estimation based on vehicle sensors and data mining," *Cogent Engineering*, vol. 5, no. 1, pp. 1449428, 2018.

[22] J.-H. Choi, and J.-S. Lee, "EmbraceNet: A robust deep learning architecture for multimodal classification," *Information Fusion*, vol. 51, pp. 259-270, 2019.

[23] U. Johnson, and A. Ivarsson, "Psychosocial factors and sport injuries: prediction, prevention and future research directions," *Current Opinion in Psychology*, vol. 16, pp. 89-92, 2017.

[24] B. Xie, M. Sidulova, and C. H. Park, "Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion," *Sensors*, vol. 21, no. 14, pp. 4913, 2021.

[25] J. Heredia, E. Lopes-Silva, Y. Cardinale, J. Diaz-Amado, I. Dongo, W. Graterol, and A. Aguilera, "Adaptive multimodal emotion detection architecture for social robots," *IEEE Access*, vol. 10, pp. 20727-20744, 2022.

[26] R. Li, and Z. Liu, "Stress detection using deep neural networks," *BMC Medical Informatics and Decision Making*, vol. 20, pp. 1-10, 2020.

[27] Q. Xu, T. L. Nwe, and C. Guan, "Cluster-based analysis for personalized stress evaluation using physiological signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 275-281, 2014.

[28] Y. Li, B. Wu, X. Li, Q. Zhou, and X. Yang, "Research on mental stress recognition of depressive disorders in patients with androgenic alopecia based on machine learning and fuzzy K-means clustering," *Frontiers in Genetics*, vol. 12, pp. 751791, 2021.

[29] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984.

[30] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18-28, 1998.

[31] P. Shermadurai, and K. Thiyagarajan, "Deep Learning Framework for Classification of Mental Stress from Multimodal Datasets," *Revued' Intelligence Artificielle*, vol. 37, no. 1, pp. 155, 2023.

[32] D. Giakoumis, D. Tzovaras, and G. Hassapis, "Subject-dependent biosignal features for increased accuracy in psychological stress detection," *International Journal of Human-Computer Studies*, vol. 71, no. 4, pp. 425-439, 2013.

[33] C. Gong, "Analysis of athletes' stadium stress source based on improved layered K-means algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 4, pp. 5905-5914, 2020.