Journal of Network Intelligence ©2025 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **10**, Number **4**, November **2025**

Identification of Psychological Stressors in Athletes Based on Improved Fuzzy C-mean Clustering

Ming-Xia Yang

Qinhuangdao Vocational and Technical College
Qinhuangdao 066100, P. R. China
qhdymx12@163.com

Zhi-Hong Hou*

Qinhuangdao Vocational and Technical College
Qinhuangdao 066100, P. R. China
houzhihong@qvc.edu.cn

Ying Lv

School of Information Technology and Engineering
St. Paul University Philippines
Tuguegarao 3500, Philippines
xa3506@163.com

*Corresponding author: Zhi-Hong Hou
Received July 18, 2024, revised January 9, 2025, accepted May 7, 2025.

Abstract. *This paper proposes a method based on improved Fuzzy C-means (FCM) clustering for identifying psychological stressors in athletes. Firstly, the characteristics and influencing factors of athletes' psychological stressors were analyzed, and the importance of psychological stress on athletes' training and competition performance was clarified. Subsequently, combined with the improved FCM clustering algorithm, the psychological stressors of athletes were effectively classified and identified. To enhance the robustness of FCM algorithm, this paper proposes an FCM algorithm based on adaptive neighbor information. Integrating the neighborhood information of sample points and the neighborhood information of class center points into the basic FCM model provides more data structure information for the clustering process, which is used to guide the cluster partitioning process in the clustering algorithm, improve the stability of the algorithm, and more accurately classify psychological stressors. The effectiveness and superiority of the proposed method have been demonstrated through experimental verification of a real athlete psychological dataset. The experimental results show that this method has high accuracy and effectiveness in identifying athletes' psychological stressors, and can provide strong support for athletes' mental health management.*
Keywords: fuzzy C-mean clustering; adaptive nearest neighbor; athletes; stressors

---

1. Introduction.

With the modern competitive sports and the increasingly fierce sports competition, currently, a clear trend has emerged in sports: in pursuit of improved athletic performance, while tapping the physiological potential of athletes, it is common to pay attention to tapping the psychological potential of athletes. This trend not only makes sport psychology come into being, but also becomes the driving force for the rapid development of sport psychology.

In the world of competitive sport, athletes face many psychological stressors that can significantly affect their performance and health. Identifying these stressors is critical to providing effective psychological support and intervention. Traditional methods of identifying psychological stressors typically rely on self-report questionnaires and interviews, which can be subjective and time-consuming [1]. To address these limitations, data-driven approaches are becoming increasingly popular, and clustering algorithms are key tools for revealing patterns and insights in psychological data.

Clustering algorithms are widely applied in the field of pattern recognition [2], image processing [3], and data mining [4], which can help people to analyze the data, understand the essential structural features of the data, and obtain useful information from the data.

One of the most widely used clustering algorithms is the Fuzzy C-mean (FCM) algorithm [5]. Unlike hard clustering methods, FCM enables data points to belong to multiple clusters with varying degrees of membership, making it well suited to deal with the ambiguity and overlap inherent in psychological data. However, traditional FCM algorithms have limitations, especially in dealing with high dimensionality and complex data distributions, which may lead to sub-optimal clustering results.

To address the aforementioned issues, this paper proposes an Adaptive Neighbour Fuzzy C-Means (ANFCM) clustering algorithm based on adaptive nearest neighbor information. The method is able to decrease the influence of data noise on the clustering effect and improve the convergence of the algorithm. The method was validated on a dataset of psychological indicators of athletes and proved its effectiveness in accurately identifying key stressors.

The results of this study provide valuable insights into the psychological stressors faced by athletes and provide a robust method for their identification. This enhanced understanding can inform the development of targeted psychological interventions and support systems, ultimately contributing to the optimization of athletes' mental health and performance.

1.1. **Related work.** When label information is unavailable, partitioning data into different segments is challenging. Thus, clustering algorithms were proposed. Clustering algorithms include K-means clustering [6], Fuzzy C-means clustering [5], and Spectral clustering [7], among others.

A common issue with FCM clustering is its sensitivity to noise and outliers [8, 9]. To address this, researchers use sparse norms to measure distance, thereby suppressing the contribution of outliers to the objective function. Chang et al. proposed a non-convex optimization model to evaluate each feature's contribution to the objective function [10]. Zhang et al. proposed an algorithm combining fuzzy clustering and spectral clustering, introducing the σ-norm to adaptively improve FCM's stability against small or large outliers [11]. Chuang et al. proposed a two-step FCM algorithm [12], and Cai et al. incorporated local spatial and local gray-level relationships to ensure noise resistance and detail preservation in images [13]. However, data often contains noise, which can affect clustering results. Inspired by the concept of shrinkage modes [14, 15], researchers perform clustering on flexible manifolds rather than in the original data space to avoid the impact of noise on data structure. To obtain the appropriate manifold structure for the original data, shrinkage mode learning is conducted. Shrinkage modes map data onto flexible manifolds of the same dimensionality, providing better noise resistance and more stable clustering [16].

Subsequently, the method of least squares approximation was used to construct constrained nonlinear programming functions, with the within-cluster sum of squares function becoming a common form of clustering objective function. Thereafter, various improvements to the FCM algorithm were made from different perspectives. Fan et al. [17] introduced a suppression parameter, modifying the membership values of FCM under constraint conditions to amplify the highest membership and suppress others, accelerating the attainment of optimal membership values. This combines the fast convergence of hard C-means clustering with the good partitioning performance of fuzzy C-means clustering. Hung et al. [18] noted that the selection of the suppression parameter in the suppression fuzzy C-means algorithm should relate to separation intensity, and combining the robustness concepts based on distance and similarity proposed by Wu and Yang [19, 20], they proposed an exponential function to calculate the suppression parameter, termed exponential separation intensity. Yang et al. [21] proposed a truncated fuzzy C-means algorithm by setting a truncation parameter, allowing membership to be 1, thereby reducing the impact of noise on the algorithm. Xiong et al. [22] proposed a hybrid algorithm combining the chaotic particle swarm algorithm with fuzzy C-means, achieving higher image segmentation accuracy and better noise robustness. Javadian et al. [23] addressed the issue that membership determination did not consider the shape and density of clusters, reprocessing the FCM algorithm's output based on cluster shape and density, and statistical hypothesis testing showed that this algorithm could improve the clustering effect of FCM.

1.2. **Motivation and contribution.** Inspired by several successful algorithmic applications of local structural information in enhancing clustering performance, this paper proposes an Adaptive Neighbour Fuzzy C-Means (ANFCM) clustering algorithm derived from adaptive nearest neighbor information. Specifically, for each sample point, according to the Euclidean distance metric between the remaining sample points and its Euclidean distance, based on the a priori assumption that the closer sample points are more likely to be the nearest neighbors, the local structural information of the data is mined to guide the clustering process, so as to achieve the effect of attenuating the influence of noise and outlier points.

First, the similarity of sample points and the similarity between cluster centers and sample points are learned through the nearest-neighbor information, and the local structural information of cluster centers and sample points is mined to guide the clustering process; second, the above two types of similarity are introduced into the traditional FCM framework to compensate for the FCM's single Euclidean distance-squared measure, which enables the algorithm to pay attention to the local neighborhood information while considering the global clustering structure to improve the algorithm's stability and reduce the sensitivity to noise and outliers. The method is validated on the athletes' psychological index dataset, which proves its effectiveness in accurately identifying key stressors.

2. **Analysis of relevant principles.**

2.1. **Fuzzy C-mean clustering algorithm.** In the Fuzzy Partition defined by Ruspini, the dataset X is divided into C fuzzy subsets $\{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_C\}$ and the affiliation μ_{ij} of the samples is extended from the set $\{0, 1\}$ to the interval $[0, 1]$ using the fuzzy set theory proposed by Zadeh in 1965, μ_{ij} subject to the satisfaction of the condition $\mu_{ij} \in E_f$, where:

$$E_f = \left\{ \mu_{ij} \mid \mu_{ij} \in [0,1]; \sum_{i=1}^{C} \mu_{ij} = 1,\ \forall j;\ 0 < \sum_{j=1}^{n} \mu_{ij} < n,\ \forall i \right\}. \tag{1}$$

This in turn extends the concept of hard C delimitation to fuzzy C delimitation, derived as follows:

$$M_f = \left\{ U \in \mathbb{R}^{C\times n} \mid \mu_{ij} \in [0,1]; \sum_{i=1}^{C} \mu_{ij} = 1,\ \forall j;\ 0 < \sum_{j=1}^{n} \mu_{ij} < n,\ \forall i \right\}. \tag{2}$$

Dunn, on the concept of fuzzy partitioning based on Ruspini's definition, generalized the objective function of hard C-mean clustering, i.e., the intraclass error sum-squared function, to a function intraclass weighted error sum-squared function. In order to ensure that this generalization is meaningful so as not to produce a nontrivial solution, thereafter, Bezdek extended Dunn's objective function to a more general form by introducing a parameter m and gave a more general description of the objective function in Equation (3):

$$J_{FCM} = \sum_{i=1}^{C} \sum_{j=1}^{n} \mu_{ij}^{m} \|x_j - p_i\|^2, \quad m > 1, \quad U \in M_f \tag{3}$$

where m is known as the fuzzy weighted index, which extends the weighted index of affiliation in the objective function J_F from a fixed value of 2 to all positive real numbers greater than 1.

The criterion of clustering is to take the minimal value of J_{FCM}. Since the samples are independent of each other, i.e., the columns in the fuzzy division matrix U are independent of each other, and then by the constraints $\sum_{i=1}^{C} \mu_{ij} = 1$ of the objective function J_{FCM} and affiliation μ_{ij}, the Lagrange multiplier method [24] can be utilized to solve for the extremes of this condition, and the coefficients λ_j $(j = 1, 2, \ldots, n)$ can be introduced to bring the constraints of each sample into the objective function:

$$L_{FCM} = \sum_{i=1}^{C} \sum_{j=1}^{n} \mu_{ij}^{m} \|x_j - p_i\|^2 + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{C} \mu_{ij} - 1 \right) \tag{4}$$

$$\frac{\partial L_{FCM}}{\partial \mu_{ij}} = m\|x_j - p_i\|^2 \mu_{ij}^{m-1} + \lambda_j = 0 \tag{5}$$

$$\mu_{ij} = \left(-\frac{\lambda_j}{m} \right)^{\frac{1}{m-1}} \frac{1}{\|x_j - p_i\|^{\frac{2}{m-1}}} \tag{6}$$

In order to solve for μ_{ij}, λ_j must be eliminated, so the constraint $\sum_{i=1}^{C} \mu_{ij} = 1$ is reused, and Equation (7) is substituted into the constraint:

$$\sum_{i=1}^{C} \mu_{ij} = \left(-\frac{\lambda_j}{m} \right)^{\frac{1}{m-1}} \sum_{i=1}^{C} \frac{1}{\|x_j - p_i\|^{\frac{2}{m-1}}} = 1 \tag{7}$$

From the above, replacing i by k for clarity, we have:

$$\left(-\frac{\lambda_j}{m} \right)^{\frac{1}{m-1}} = \frac{1}{\sum_{i=1}^{C} \frac{1}{\|x_j - p_i\|^{\frac{2}{m-1}}}} = \frac{1}{\sum_{k=1}^{C} \frac{1}{\|x_j - p_k\|^{\frac{2}{m-1}}}} \tag{8}$$

Substituting Equation (9) into Equation (7) yields an expression for μ_{ij}:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_j - p_i\|}{\|x_j - p_k\|} \right)^{\frac{2}{m-1}}} \tag{9}$$

By combining Equations (9) and (10), we can obtain the standard membership update rule for the Fuzzy C-means clustering algorithm.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_j - p_i\|}{\|x_j - p_k\|} \right)^{\frac{2}{m-1}}}, \quad H_j = \varnothing; \quad \mu_{ij} = 0,\ \forall i \in \bar{H}_j;\ \mu_{ij} = 1,\ i \in H_j,\ H_j \neq \varnothing \tag{10}$$

The LFCM is obtained by taking the partial derivative of p_i:

$$\sum_{j=1}^{n} \mu_{ij}^{m} [-2l(x_j - p_i)] = 0 \tag{11}$$

Final expression:

$$p_i = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} x_j}{\sum_{j=1}^{n} \mu_{ij}^{m}}, \quad i = 1, 2, \dots, C \tag{12}$$

From the above, it is not difficult to find that μ_{ij} and p_i are related to each other, so as long as the dataset X, the number of categories C, and the fuzzy weighting index m are known, and then given the initial clustering center matrix $P(0)$ or initial fuzzy partitioning matrix $U(0)$, the clustering result is obtained through computation [25].

2.2. **Robust sparse fuzzy K-mean clustering.** In order to enhance the stability of FCM against outliers, the $\ell_{2,1}$ parameter and the truncated ℓ_1 parameter are used to replace the original square-parameter to reduce the influence of outliers on the clustering results. The Robust Sparse Fuzzy K-Mean (RSFKM) model for $\ell_{2,1}$ norms clustering is defined as:

$$\min_{U,M} \sum_{i=1}^{m} \sum_{k=1}^{C} u_{ik} \|x_i - m_k\|_2 + \gamma \|U\|_F^2 \tag{13}$$

The RSFKM model for the truncated ℓ_1 norm is defined as:

$$\min_{U,M} \sum_{i=1}^{n} \sum_{k=1}^{C} u_{ik} \min(\|x_i - m_k\|_2, \varepsilon) + \gamma \|U\|_F^2 \tag{14}$$

where if $\gamma = 0$, the affiliation vector for each sample point is sparse (only one element is non-zero, all other elements are zero).

When $\gamma > 0$, the affiliation vector is denser than when $\gamma = 0$. By adjusting γ, the sparsity of the affiliation vector is an asymptotic change; as γ gradually increases, the affiliation vector contains more and more non-zero elements. When γ reaches a large value, all the elements of the affiliation vector are non-zero, and then the affiliation vector is non-sparse. A reasonable sparsity of the affiliation vector can be found by parameter optimization to obtain more accurate clustering results.

2.3. **Psychological stressors in athletes.** Athletes' psychological stressors are factors that may cause athletes to feel tension, anxiety, stress, and other adverse emotions during training and competition. These psychological stressors can be multifaceted and mainly include the following aspects:

1. **Competition performance pressure:** Athletes demand too much from their own performance and results, fearing failure or losing the goals set before the competition, leading to tension and anxiety.
2. **Social pressure:** Expectations and evaluations from social environments such as family, team, coaches, media, and spectators create psychological pressure on athletes.

3. **Self-expectation and self-evaluation:** Athletes have certain expectations of their own performance and results. Excessively high or low self-expectation and self-evaluation will increase psychological pressure.
4. **Lack of self-confidence and self-doubt:** Athletes with low self-confidence, self-doubt or doubt in their own abilities are vulnerable to psychological stress.
5. **Fear and anxiety:** Fear and anxiety about various new situations, difficulties and challenges that may be faced during the competition generate stress.
6. **Excessive training load:** Training is too intense and frequent, resulting in physical and mental fatigue and affecting the athlete's psychological state.
7. **Injury, illness and physical condition:** Athletes' physical injuries or poor physical condition affect their performance and mindset, resulting in psychological stress.
8. **Team cooperation and interpersonal relationship:** Poor cooperation within the team, interpersonal relationship problems or conflicts among teammates will bring about psychological pressure on athletes.

Athletes' psychological stressors are multifaceted and come from a variety of internal and external factors. Recognizing and effectively managing these psychological stressors is crucial to improving athletes' psychological quality and competitive performance.

3. Adaptive nearest neighbor information fuzzy C mean clustering.

3.1. Modeling.

Each data point can be regarded as a nearest neighbor of other data points, but the similarity between different data points is different, and based on the premise assumption that data points with smaller distances should have a greater probability of becoming neighbors, it can be assumed that the rate of sample points with closer distances belonging to the same category is larger. Therefore, the adaptive nearest neighbor information will adaptively select the most relevant nearest neighbor points for category classification based on the similarity of each data point with other data points.

Euclidean distance is used as a measure of distance. Given a dataset $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d\times n}$, where d is the dimension and n is the number of sample points in the dataset. $x_j \in \mathbb{R}^d$ is the j-th sample point, and the probability of the rest of the sample points becoming neighbors with x_j is set to s_{jk}, with a smaller distance $\|x_j - x_k\|_2^2$ corresponding to a larger near-neighbor probability of s_{jk}. The information about the near-neighborhoods is embodied in the form of the similarity matrix S, and the element s_{jk} in the similarity matrix of the above-mentioned dataset X between the sample points, i.e. the near-neighborhood probability between the two sample points of the n sample points, is the probability of the near-neighborhood between the two sample points. The similarity matrix of each sample point in the above data set X, each element s_{jk}, is the probability of closer neighbor between two n sample points. The similarity matrix $X \in \mathbb{R}^{n\times n}$ of data set X can be solved as follow:

$$\min_{s^i,\, 0\le s_{jk}\le 1} \sum_{k=1}^{n} \left\|x_j - x_k\right\|_2^2 s_{jk} + \lambda\, s_{jk}^2 \tag{15}$$

This paper defines the neighborhood information of a sample point x_i as G_{xy}. G_{xy} is a value, and the neighborhood information of all sample points in the dataset X forms a vector $G_x \in \mathbb{R}^n$, where G_{xy} is its j-th element. The regularization parameter λ controls the sparsity of the similarity matrix S. The definition of G_{xy} is as follows:

$$G_{x_j} = \min_{\boldsymbol{s}} \sum_{k=1}^{n} \|x_j - x_k\|_2^2\, s_{jk} + \lambda s_{jk}^2 \quad \text{s.t.} \quad s_{jk} \ge 0,\ \sum_{k=1}^{n} s_{jk} = 1 \tag{16}$$

Assuming that the n sample points in the dataset X can be classified into c categories, the i-th column vector v_i in the class center matrix $V \in \mathbb{R}^{d\times c}$ is the center point of the i-th category. Similarly, the proximity information between the class center point v_i and the n sample points is defined as G_{v_i}. The definition of G_{v_i} is:

$$G_{v_i} = \min_{\boldsymbol{s}'} \sum_{k=1}^{n} \|v_i - x_k\|_2^2 \; s'_{jk} + \lambda s'^2_{jk} \quad \text{s.t.} \quad s'_{jk} \geq 0, \; \sum_{k=1}^{n} s'_{jk} = 1 \tag{17}$$

After obtaining the nearest-neighbor information G_x of the sample points and G_v of the class centers, we incorporate them into the basic FCM to obtain the ANFCM model:

$$\min_{U,V} J = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \left(\|x_j - v_i\|_2^2 + \alpha\big(G_{x_j} - G_{v_i}\big)^2 \right) \tag{18}$$

3.2. **Model optimization.** The ANFCM model considered in this paper is specifically written as:

$$\begin{cases} \min\limits_{U,V} J = \sum\limits_{i=1}^{c}\sum\limits_{j=1}^{n} u_{ij}^m \left(\|x_j - v_i\|_2^2 + \alpha\big(G_{x_j} - G_{v_i}\big)^2 \right), \\ G_{x_j} = \min\limits_{\boldsymbol{s}} \sum\limits_{k=1}^{N} \|x_j - x_k\|_2^2 \; s_{jk} + \lambda s_{jk}^2, \\ G_{v_i} = \min\limits_{\boldsymbol{s}'} \sum\limits_{k=1}^{N} \|v_i - x_k\|_2^2 \; s'_{jk} + \lambda s'^2_{jk}. \end{cases} \tag{19}$$

The model contains 3 minima optimization problems. Since the value of G_{xj} is only related to the data X, it can be solved at the beginning. Define first:

$$d_{jk} = \|x_j - x_k\|_2^2 . \tag{20}$$

$$G_x = \min_{\mathbf{S}} \sum_{j=1}^{N}\sum_{k=1}^{N} s_{jk}\, d_{jk} + \lambda \|\mathbf{S}\|_F^2 \tag{21}$$

$$= \min_{\{\mathbf{s}_j\}_{j=1}^N} \sum_{j=1}^{N} \big(\mathbf{s}_j^\mathsf{T}\mathbf{d}_j + \lambda\|\mathbf{s}_j\|_2^2\big) = \min_{\{\mathbf{s}_j\}_{j=1}^N} \sum_{j=1}^{N} \left\| \mathbf{s}_j + \frac{\mathbf{d}_j}{2\lambda} \right\|_2^2 - \frac{1}{4\lambda^2}\sum_{j=1}^{N} \mathbf{d}_j^\mathsf{T}\mathbf{d}_j \tag{22}$$

$$\iff \min_{\{\mathbf{s}_j\}_{j=1}^N} \sum_{j=1}^{N} \left\| \mathbf{s}_j + \frac{\mathbf{d}_j}{2\lambda} \right\|_2^2 \tag{23}$$

For this optimization problem, it can be solved based on the Lagrangian method and the KKT condition:

$$L(\mathbf{s}_j, \eta, \boldsymbol{\beta}_j) = \frac{1}{2}\left\| \mathbf{s}_j + \frac{\mathbf{d}_j}{2\lambda} \right\|_2^2 - \eta\big(\mathbf{s}_j^\mathsf{T}\mathbf{1} - 1\big) - \boldsymbol{\beta}_j^\mathsf{T}\mathbf{s}_j \tag{24}$$

the solution steps are:

$$\lambda_j = \frac{k}{2}\, d_{j,k+1} - \frac{1}{2}\sum_{i=1}^{k} d_{ji} \tag{25}$$

$$\lambda = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{k}{2}\, d_{j,k+1} - \frac{1}{2}\sum_{i=1}^{k} d_{ji} \right) \tag{26}$$

Algorithm 1 Solving for sample point nearest neighbor information G_x

1: **Input:** Data matrix $X \in \mathbb{R}^{n\times n}$, class clusters c, adaptive nearest neighbors k
2: **Output:** Adaptive nearest neighbor information vector $G_x \in \mathbb{R}^{l\times n}$
3: **Begin**
4: Calculate the distance matrix $D \in \mathbb{R}^{n\times n}$;
5: Based on the given number of adaptive nearest neighbors k, the parameter λ;
6: Calculating Lagrange multipliers;
7: Similarity matrix $S \in \mathbb{R}^{n\times n}$;
8: Calculate the nearest neighbor information of the sample point x_j;
9: **Return** Nearest neighbor information vector G_x

$$\eta = \frac{1}{k} + \frac{1}{2k\,\lambda_j}\sum_{i=1}^{k} d_{ji} \tag{27}$$

$$s_{ji} = \left(-\frac{d_{ji}}{2\lambda_j} + \eta\right)_+ \tag{28}$$

Organize the above derivation into Algorithm 1 for solving the G_{xj} problem.

The algorithm starts with a random initialization of the affiliation matrix U and the class centroid matrix V. The optimization of G_{v_i} is related to the data X and the class centroid matrix V. Each time, based on the updated class centroid matrix V, the solution process of G_{v_i} is similar to that of G_{xj}, and its minimum value is calculated using Algorithm 2.

Algorithm 2 Class centroid nearest-neighbor information G_v solving

1: **Input:** Data matrix $X \in \mathbb{R}^{d\times n}$, clustering prototype matrix $V \in \mathbb{R}^{d\times c}$, clusters c, adaptive nearest neighbors k
2: **Output:** Class centroid adaptive nearest neighbor information vector
3: **Begin**
4: Calculate distance matrix;
5: Calculated parameters λ;
6: Calculating Lagrange multipliers;
7: Similarity matrix $S \in \mathbb{R}^{c\times n}$;
8: Compute the nearest-neighbor information for the class centroid v_i;
9: **Return** Nearest neighbor information vector G_v

When optimizing the objective function J, G_x and G_{v_i} are treated as constants and solved according to the Lagrange multiplier method. Construct the Lagrange function:

$$L = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\big(\|x_j - v_i\|_2^2 + \alpha\,(G_x - G_{v_i})^2\big) + \sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{c} u_{ij} - 1\right). \tag{29}$$

Let the function L take the partial derivative with respect to u_{ij} and let it be zero:

$$\frac{\partial L}{\partial u_{ij}} = m\big(\|x_j - v_i\|_2^2 + \alpha(G_x - G_{v_i})^2\big)\,u_{ij}^{m-1} + \lambda_j = 0 \tag{30}$$

$$u_{ij}^{m-1} = -\frac{\lambda_j}{m\big(\|x_j - v_i\|_2^2 + \alpha(G_x - G_{v_i})^2\big)} \tag{31}$$

$$u_{ij} = \left(\frac{\lambda_j}{m}\right)^{\frac{1}{m-1}} \left(\|x_j - v_i\|_2^2 + \alpha(G_x - G_{v_i})^2\right)^{-\frac{1}{m-1}}. \tag{32}$$

Combined with the constraint $\sum_{i=1}^{c} u_{ij} = 1$, this gives:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_j - v_i\|_2^2 + \alpha(G_x - G_{v_i})^2}{\|x_j - v_k\|_2^2 + \alpha(G_x - G_{v_k})^2}\right)^{\frac{1}{m-1}}}. \tag{33}$$

Let the function L take the partial derivative with respect to v_i and let it be zero:

$$\frac{\partial L}{\partial v_i} = \sum_{j=1}^{n}\left(-2u_{ij}^m (x_j - v_i)\right) = 0 \tag{34}$$

$$-2\sum_{j=1}^{n} u_{ij}^m x_j + 2\sum_{j=1}^{n} u_{ij}^m v_i = 0 \tag{35}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}. \tag{36}$$

The ANFCM algorithm flow is shown in Figure 1, and the overall solution steps are summarized in Algorithm 3.

Algorithm 3 ANFCM Iterative Solution Method

```
Input: Data matrix X ∈ R^{d×n}, class clusters c, nearest neighbors of sample points
  and class centers k_x, k_v
Output: Affiliation matrix U
Begin
Randomly initialize the affiliation matrix U;
Initialize the clustering prototype matrix V;
Compute sample-point nearest neighbor information G_X;
while U not converge do
    Update class-centroid nearest neighbor information G_V;
    Update the affiliation matrix U;
    Update the clustering prototype matrix V;
end while
Return Nearest neighbor information vector G_V
```

4. Experiment.

4.1. Datasets.

In the comparative experiment, we used four datasets [26, 27], as shown in Table 1, which are commonly used datasets for measuring athletes' psychological stressors.

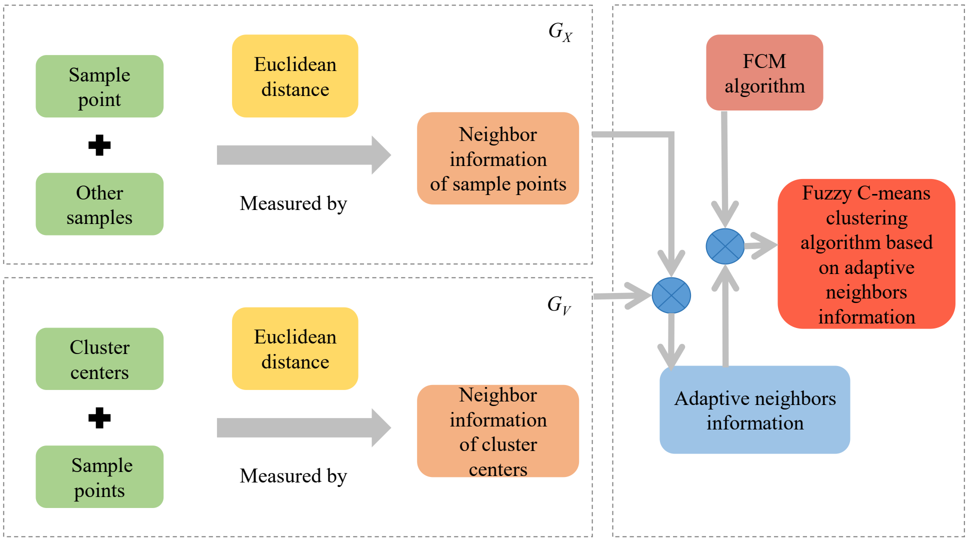

Figure 1. Flowchart of ANFCM algorithm

Table 1. Description of benchmark datasets

Dataset	Number of Samples	Characteristic number	Number of categories
APSQ	351	34	2
SAS	373	4	2
ABQ	359	64	2
CSAI-2	106	7	2

4.2. **Comparison algorithm.** This article uses four advanced related clustering methods, KM, FCM, Fuzzy Closeness and Separation Clustering (FCS) algorithm, and Aggregation Fuzzy K-means Clustering (AFKM) algorithm, as comparative algorithms.

1. K-Means is a prominent clustering algorithm that partitions a dataset into multiple clusters, ensuring each data point is assigned to exactly one cluster.
2. The FCM algorithm extends K-Means by treating each cluster as a fuzzy set, with a membership function indicating the likelihood of each data point belonging to a cluster.
3. FCS allows both hard and fuzzy memberships to coexist in the clustering results.
4. Compared with FCM, the AFKM algorithm uses regularization parameters to adjust fuzzy membership degrees and introduces maximum entropy information to optimize clustering partitions.

4.3. **Parameter settings.** In FCM–type algorithms, a large number of experiments have shown that setting the membership weight index to 2 can achieve better results. Based on experience, the membership weight index of all algorithms in this section is set to 2, and the remaining parameters that need to be adjusted are selected using a grid–search strategy. The ANFCM algorithm has three parameters that need to be adjusted. The first parameter is the regularization parameter α, which is used to adjust the degree of influence of global information and sample–point neighbor information on clustering. The values are $\{0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$. The second parameter is the count of nearest–neighbor sample points k_x, and the third parameter is the count of nearest–neighbor clustering prototypes k_v. The settings for both parameters are the same, $\{2, 3, 5, 7, 9, 15, 19, 25\}$. The regularization parameter λ of the AFKM algorithm is set to $\{0.001, 0.01, 0.1, 1, 2, 5, 10, 100\}$.

4.4. **Evaluating indicator.** ACC, NMI, and Purity are commonly used indicators to evaluate the quality of clustering results. Among them, ACC is the accuracy of clustering, which measures the extent of matching between clustering results and real labels; NMI is a standardized mutual information measure that measures the similarity between clustering results and real labels. Its values range from 0 to 1, with higher values indicating that clustering results are more similar to real labels; Purity is purity, which measures the proportion of data points in the same category in the clustering results. It calculates the true labels that appear the most frequently in each cluster, adds up the number of times these labels appear, and divides by the total number of data points to obtain the purity of the clustering result. The value ranges from 0 to 1, with higher values indicating more data points in the same category in the clustering result.

4.5. **Comparative experimental results and analysis.** Due to the impact of random initialization on clustering results, all experimental results were obtained by taking the average of 10 randomly initialized clustering centers under the same conditions. Each clustering algorithm performs clustering on four benchmark datasets, with evaluation metrics shown in Figure 2, Figure 3, and Figure 4, respectively. Blue represents the ANFCM algorithm proposed in this paper, green represents the KM algorithm, yellow represents the FCM algorithm, red represents the FCS algorithm, and black represents the AFKM algorithm. From Figure 4, it is clear that the clustering performance of ANFCM algorithm on APSQ and SAS datasets is more than 10% higher than the indicators of the compared algorithms; Overall, compared with traditional FCM, the ANFCM algorithm showed an average improvement of 11.8725%, 15.4425%, and 6.6163% in the Accu Race, NMI, and Purity metrics across all datasets in the experiment. The introduction of local neighbor information effectively guides the learning of membership matrices, thereby improving the accuracy of clustering algorithms.

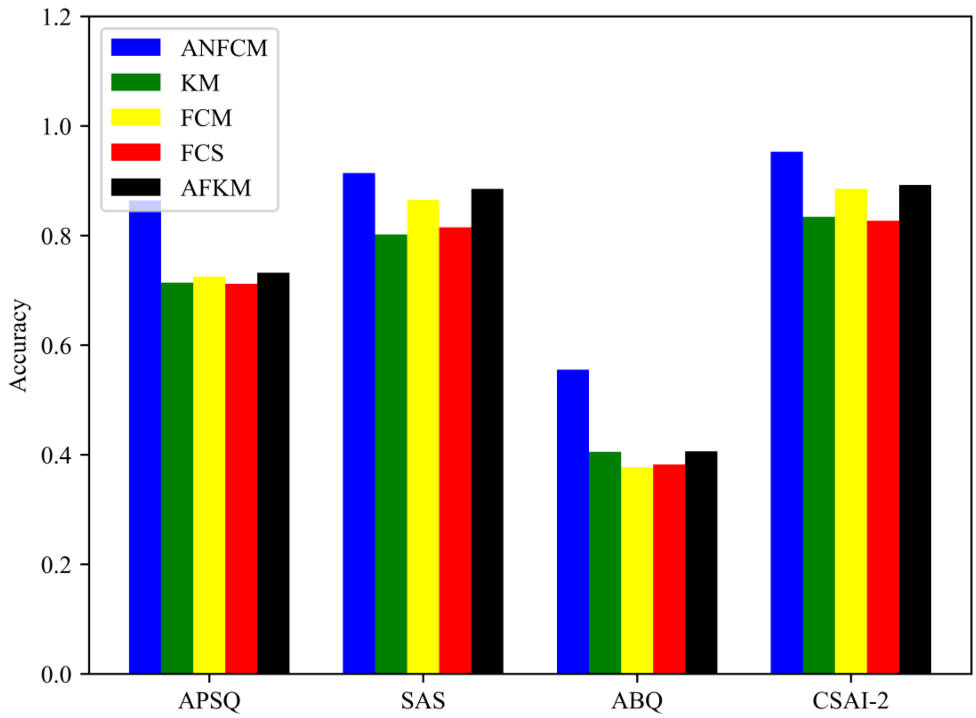

Figure 2. Accuracy values for each algorithm on 4 datasets

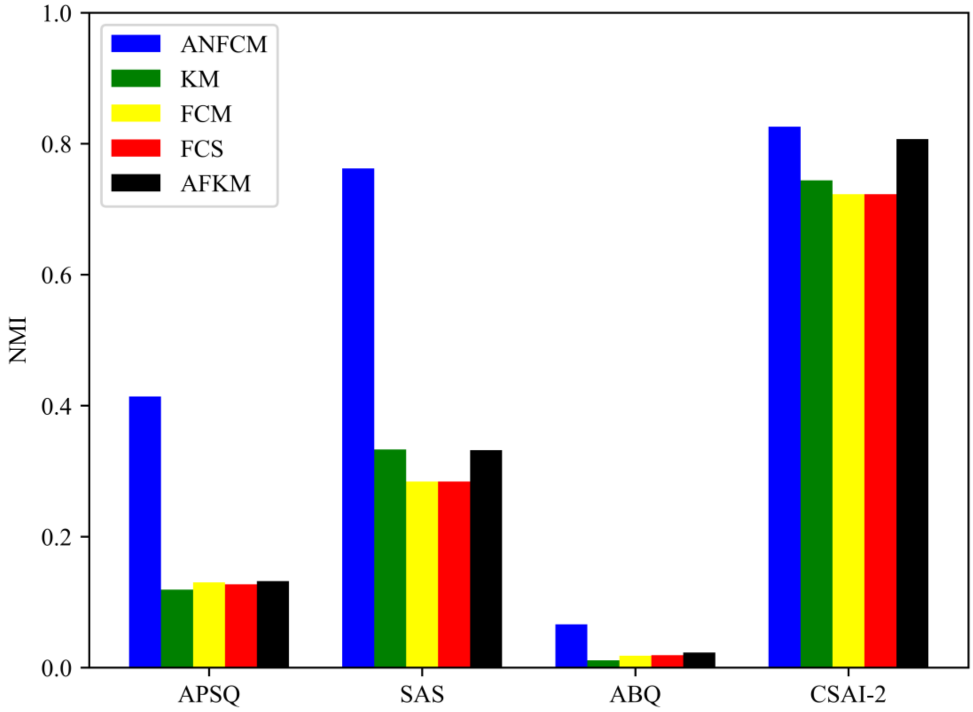

Figure 3. NMI values for each algorithm on 4 datasets

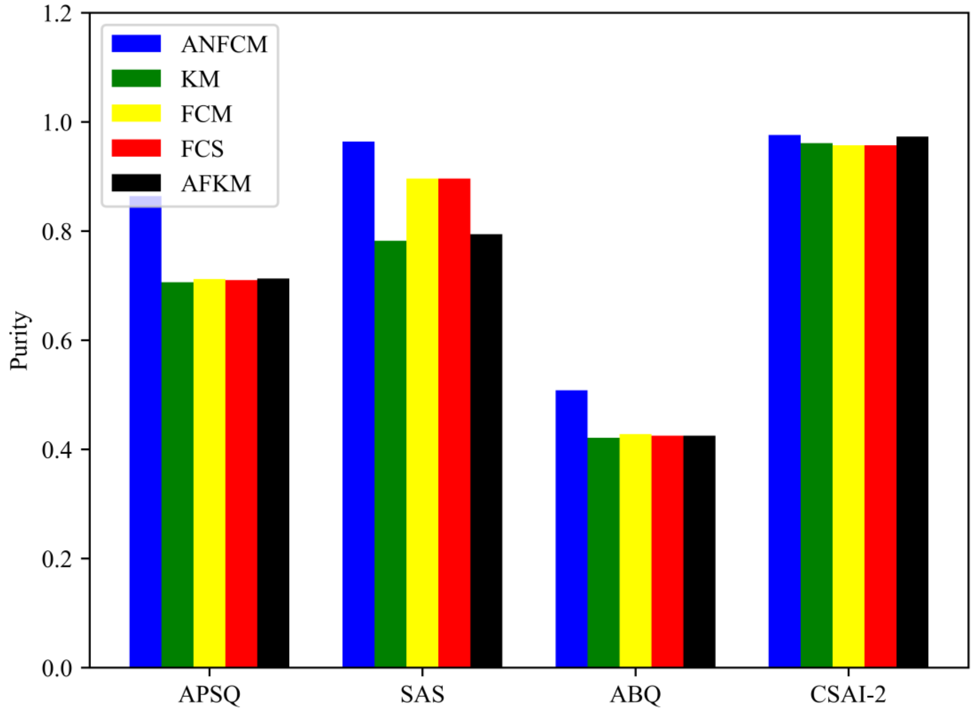

Figure 4. Purity values for each algorithm on 4 datasets

5. **Conclusions.** This article proposes a fuzzy C-means clustering algorithm that simultaneously mines the local structural information of cluster centers and sample points to guide the clustering process. While ensuring the advantages of the fuzzy C-means algorithm, it reduces the influence of noise and outliers, and improves the stability of the

algorithm. Although traditional FCM algorithms are useful, they often struggle to handle high-dimensional and complex data, resulting in suboptimal clustering results. Our improved FCM algorithm effectively addresses these challenges.

The enhanced algorithm was validated using a dataset of athlete psychological indicators, demonstrating its superior performance in identifying key psychological stressors. The results indicate that the improved FCM algorithm not only provides more precise and stable clustering results, but also provides deeper insights into specific stressors that affect athletes.

These findings have significant implications for the field of sports psychology as they provide a powerful, data-driven approach for identifying sources of psychological stress. This method can be used to provide information for targeted psychological interventions and support strategies, ultimately contributing to the mental health and performance improvement of athletes. Future research can explore the application of this improved clustering method in other fields within and outside of psychology, further demonstrating its universality and effectiveness.

REFERENCES

[1] R. Barić, "Psychological pressure and athletes' perception of motivational climate in team sports," *Review of Psychology*, vol. 18, no. 1, pp. 45–49, 2011.

[2] P. Melin and O. Castillo, "A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition," *Applied Soft Computing*, vol. 21, pp. 568–577, 2014.

[3] J. Yang and J. Zhu, "Cluster identification using image processing," *Particuology*, vol. 23, pp. 16–24, 2015.

[4] P.-N. Tan, M. Steinbach, and V. Kumar, "Data mining cluster analysis: basic concepts and algorithms," *Introduction to Data Mining*, vol. 487, p. 533, 2013.

[5] J. Yu and M.-S. Yang, "Optimality test for generalized FCM and its application to parameter selection," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 1, pp. 164–176, 2005.

[6] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 3, pp. 433–439, 1999.

[7] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, pp. 395–416, 2007.

[8] J. Yu and M.-S. Yang, "Optimality test for generalized FCM and its application to parameter selection," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 1, pp. 164–176, 2005.

[9] J. Yu and M.-S. Yang, "A generalized fuzzy clustering regularization model with optimality tests and model complexity analysis," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 5, pp. 904–915, 2007.

[10] X. Chang, Q. Wang, Y. Liu, and Y. Wang, "Sparse regularization in fuzzy c-means for high-dimensional data clustering," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2616–2627, 2016.

[11] R. Zhang, F. Nie, M. Guo, X. Wei, and X. Li, "Joint learning of fuzzy k-means and nonnegative spectral clustering with side information," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2152–2162, 2018.

[12] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.

[13] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.

[14] C. Hou, F. Nie, Y. Jiao, C. Zhang, and Y. Wu, "Learning a subspace for clustering via pattern shrinking," *Information Processing & Management*, vol. 49, no. 4, pp. 871–883, 2013.

[15] Z. Kang, H. Pan, S. C. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1833–1843, 2019.

[16] X. Zhao, F. Nie, R. Wang, and X. Li, "Robust fuzzy k-means clustering with shrunk patterns learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 3001–3013, 2021.

[17] J.-L. Fan, W.-Z. Zhen, and W.-X. Xie, "Suppressed fuzzy c-means clustering algorithm," *Pattern Recognition Letters*, vol. 24, no. 9–10, pp. 1607–1612, 2003.

[18] W.-L. Hung, M.-S. Yang, and D.-H. Chen, "Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 424–438, 2006.

[19] K.-L. Wu and M.-S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognition*, vol. 35, no. 10, pp. 2267–2278, 2002.

[20] M.-S. Yang and K.-L. Wu, "A similarity-based robust clustering method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 434–448, 2004.

[21] M.-S. Yang, K.-L. Wu, J.-N. Hsieh, and J. Yu, "Alpha-cut implemented fuzzy clustering algorithms and switching regressions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 3, pp. 588–603, 2008.

[22] L. Xiong, G. Tang, Y.-C. Chen, Y.-X. Hu, and R.-S. Chen, "Color disease spot image segmentation algorithm based on chaotic particle swarm optimization and FCM," *The Journal of Supercomputing*, vol. 76, pp. 8756–8770, 2020.

[23] M. Javadian, R. Vaziri, S. Haghzad Klidbary, and A. Malekzadeh, "Refining membership degrees obtained from fuzzy C-means by re-fuzzification," *Iranian Journal of Fuzzy Systems*, vol. 17, no. 4, pp. 85–104, 2020.

[24] T.-Y. Wu, J. C.-W. Lin, U. Yun, C.-H. Chen, G. Srivastava, and X. Lv, "An efficient algorithm for fuzzy frequent itemset mining," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 5787–5797, 2020.

[25] T.-Y. Wu, J. C.-W. Lin, Y. Zhang, and C.-H. Chen, "A Grid-Based Swarm Intelligence Algorithm for Privacy-Preserving Data Mining," *Applied Sciences*, vol. 9, no. 4, p. 774, 2019.

[26] J. Gao, H. Zou, F. Zhang, and T. Y. Wu, "An intelligent stage light-based actor identification and positioning system," *International Journal of Information and Computer Security*, vol. 18, no. 1/2, p. 204, 2022.

[27] W. Pedrycz, "Federated FCM: Clustering under privacy requirements," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 8, pp. 3384–3388, 2021.