



# Fuzzy Granulation for Feature Extraction in EEG-Based Stress Pattern Recognition

Sandra Eugenia Barajas-Montiel<sup>(✉)</sup> , Carlos Alberto Reyes-García ,  
and Luis Villaseñor-Pineda 

Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro No. 1, Sta. María Tonantzintla, Puebla, Mexico  
{sandybarajas,kargaxxi,villasen}@inaoep.mx

**Abstract.** The study of different techniques for the characterization of biosignals tends to facilitate and promote improvement in the results of their analysis and classification. We propose electroencephalogram (EEG) characterization through fuzzy methodologies based on fuzzy partitions. The present work explores four different fuzzy granulation techniques to extract characteristics from EEG signals to perform multi-class stress recognition. Our experiments show that feature extraction through fuzzy granulation allows to achieve results that exceed the accuracy percentage in the classification of three different stress patterns previously reached in other works focused on the same classification problem with the same database.

**Keywords:** Biosignals · Fuzzy granulation · EEG · Feature extraction · Classification · Stress patterns

## 1 Introduction

Fuzzy granulation provides a new approach to data analysis, granulation allows to set different levels of detail in the description of an object and human reasoning purposes this type of description through concepts such as small, medium, far, which are fuzzy concepts. Fuzzy granulation is based on how human reasoning describes and manipulates information [1]. It has been used for example to build information granules that describe data structures [2], to reduce redundant attributes [3], also to model time series [4]. Hence, we propose to experiment granulation of electroencephalogram (EEG) signals to extract features and recognize different stress patterns.

The reviewed literature, e.g. [5–11], demonstrates that when working with biosignals as EEG, a significant number of processing steps are required to extract features. Most related works are based on frequency domain features. This work aims to take advantage of fuzzy logic descriptive nature to characterize signals and then recognize patterns of more than two stress scenarios. No

---

This work is funded by CONACyT under grant 49245.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024  
Y. Hernández Heredia et al. (Eds.): IWAIIPR 2023, LNCS 14335, pp. 60–71, 2024.  
[https://doi.org/10.1007/978-3-031-49552-6\\_6](https://doi.org/10.1007/978-3-031-49552-6_6)

work has been found in which EEG are characterized through fuzzy granulation and then processed by a classifier for the distinction of multiple classes in EEG stress signals.

We have explored four fuzzy granulation methods based on [12] and [13], to extract features from EEG signals. To show the advantage on how fuzzy granulation is useful to represent changes in time series, we propose to adapt this approach as a feature extraction method for EEG signals. The aim was to observe the EEG signal classification results obtained after applying four fuzzy granulation techniques and determine which of them yielded the best results in multi-class stress recognition. The rest of this paper is organized as follows: Basic concepts about fuzzy granulation and the applied algorithms are explained in Sect. 2. Then, in Sect. 3 we describe experiments and results. Finally in Sect. 4 we present our conclusions.

## 2 Fuzzy Granulation

In general, when we break down an object, problem, or concept into simpler parts, it is easier to represent the object, to solve the problem, or to understand the concept. Fuzzy granulation techniques could be applied in order to discover a family of information granules to characterize the input space [14] of a classification task, as they allow the decomposition of a whole into parts. Fuzzy granulation represents an analogy to how human reasoning tends to describe entities with terms that do not have precise limits: close, far, almost, enough, etc., this type of decomposition is useful to represent changes in time series. Other applications of fuzzy granulation include the reduction of attributes as fuzzy granulation consists of extracting data from information abstraction in such a way that redundant information is eliminated, breaking down a complex problem into simpler parts/granules. The approaches followed in this work for the extraction of features in EEG signals through fuzzy granulation are presented below.

### 2.1 Fuzzy Information Granulation for Time Series

According to [4] time series granulation can be seen as a process composed by four layers: discretization, granulation, linguistic description and prediction. We base our feature extraction strategy in the first three layers of the mentioned framework. For granulation and linguistic description we adapted the methodology presented in [12]:

1. Discretization. This layer is responsible for dividing the time series into windows of the same size. Having a time series  $T = \{t_1, t_2, \dots, t_{n-1}, t_n\}$  and  $1 \leq l \leq n$  the size of the window, if  $l = n$  then the time series is represented by one window, if  $l = 1$  then  $T$  is discretized in  $n$  windows. Once  $l$  has been set, time series  $T$  can be split into windows  $W_1 = \{t_1, t_2, \dots, t_l\}$ ,  $W_2 = \{t_{l+1}, t_{l+2}, \dots, t_{2l}\}$ , ...,  $W_l = \{t_{n-l+1}, \dots, t_{n-1}, t_n\}$ .

2. Granulation. This process consists in extracting granules of the segmented windows from the previous layer they are distributed throughout the time window. Granules are a representation of the data in windows using intervals, rough sets, or other type of sets. The interest of this proposal is in using fuzzy sets for extracting granules since they provide information with different level of detail of problems with imprecise information. Fuzzy membership functions (Mf) are applied to granulate the values in each window.
3. Linguistic description. In this layer, each granule is associated with a linguistic term that describes it. In fuzzy logic, linguistic descriptors are used to represent concepts to be manipulated. When selecting a linguistic descriptor, it must be taken into account that the objective is to represent the concept to be treated in the most appropriate way. For this, different strategies can be followed such as the definition of linguistic descriptors using adjectives that describe physical characteristics. For example, to represent temperature you can use language descriptors such as cold, warm, hot. A time series could be granulated using linguistic terms as high amplitude, medium amplitude and small amplitude, each of them calculated for each window  $W_i$ .

To granulate the values in each subset, four different approaches were tested: Fuzzy Clustering, Triangular, Trapezoidal and Minmax-based granulation.

**Fuzzy Clustering.** Clustering is a grouping technique, generally applied to unsupervised learning, where given a set of objects, each object is assigned to a subset (cluster/class) of these objects, represented by a centroid. In Fuzzy Clustering, objects belong to more than one cluster with different degrees of membership in the interval [0,1]. A simple version of the Fuzzy Clustering algorithm used for granulation is presented in Table 1.

**Table 1.** Fuzzy C Means Basic Algorithm.

---

Choose a number of clusters $k$ and set $m$ the fuzzyness coefficient
Randomly assign to each object $x$ a membership degree to clusters, $W_0$
While change in $W_i$ is less than a set threshold $e$ :
Calculate the centroid for each group
$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}$
Calculate new $W_i$
For each object, calculate its belonging to the clusters and
its coefficients of being in the groups
$w_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\ x_i - c_j\ }{\ x_i - c_k\ } \right)^{\frac{2}{m-1}}}$
End while
Return c and M

---

The centroid of a cluster represents the relevant information of the granules of the segmented windows with the highest degree of membership in that cluster. Keeping only the centroid of a cluster eliminates redundant information. Clustering allows us to describe a data set in a two-level structure, first by calculating several information granules at the same time, and then describing each cluster with the resulting information granules [15]. After applying the Fuzzy C Means clustering algorithm, the centroids of each cluster are ordered. The smallest cluster is labeled as *low*, the middle one as *mid*, and the largest as *high* as shown in Fig. 1.

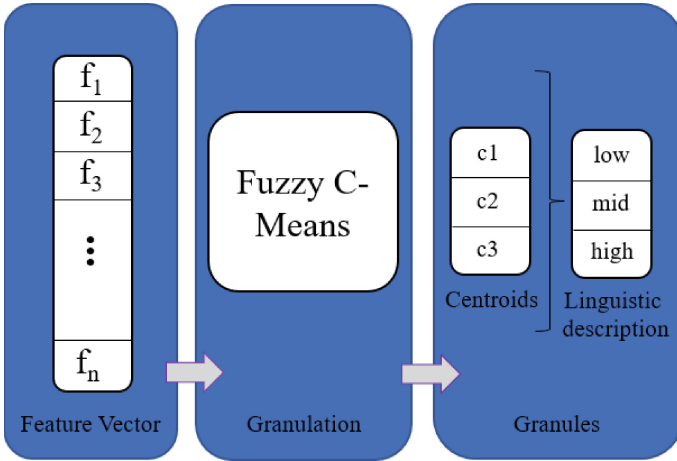


Fig. 1. Fuzzy granulation applying fuzzy clustering.

**Triangular, Trapezoidal and Minmax Granulation.** To extract features from the EEG signals each sample is sorted resulting in the time series  $T = \{t_1, t_2, \dots, t_{n-1}, t_n\}$  then they are divided in two subseries  $T_l = \{t_1, t_2, \dots, t_{n/2}\}$  and  $T_h = \{t_{(n/2)+1}, \dots, t_{n-1}, t_n\}$ .

Granules *low* and *high* are extracted form  $T_l$  and  $T_h$  respectively. The median of ordered time series  $T$  is computed to represent *mid* granule. This process is illustrated in Fig. 2. The functions used to extract granules were: Triangular-granulation (TG), Trapezoidal-granulation (TZG) and Minmax-granulation (mMG).

$$TG = \begin{cases} low = \frac{2 \sum_{j=1}^{n/2} t_j}{n/2} - median\{t_1, \dots, t_n\} \\ mid = median\{t_1, t_2, \dots, t_n\} \\ high = \frac{2 \sum_{j=n/2+1}^n t_j}{n/2} - median\{t_1, \dots, t_n\} \end{cases} \quad (1)$$

$$TZG = \begin{cases} low = \frac{2 \sum_{j=1}^{n/2} t_j}{n/2} - t_{n/2} \\ mid = median\{t_1, t_2, \dots, t_n\} \\ high = \frac{2 \sum_{j=n/2+1}^n t_j}{n/2} - t_{n/2+1} \end{cases} \quad (2)$$

$$mMG = \begin{cases} low = t_1 \\ mid = median\{t_1, t_2, \dots, t_n\} \\ high = t_n \end{cases} \quad (3)$$

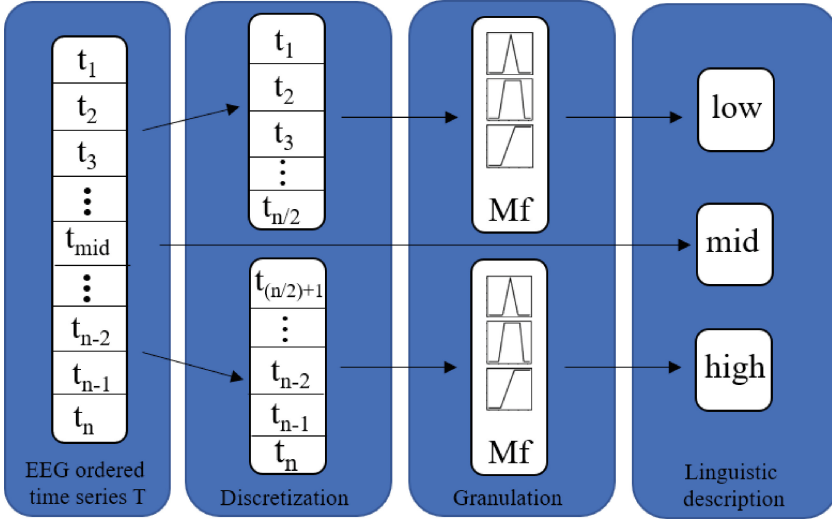


Fig. 2. Fuzzy granulation for time series.

### 3 Experiments and Results

For the development of this work, fuzzy information granulation for time series was applied, taking each channel of the EEG signal and computing granules using Fuzzy Clustering, Triangular, Trapezoidal and Minmax-based granulation.

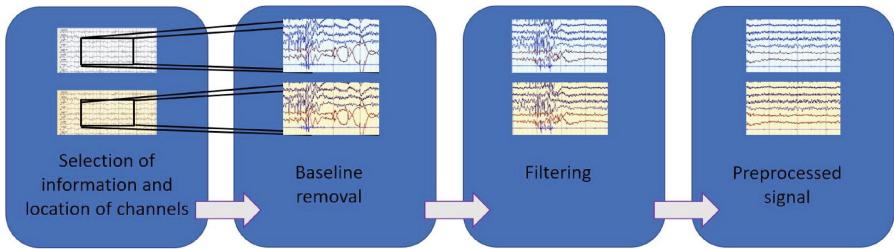
We worked with a corpus of electroencephalographic signals from 12 participants under different sound stimuli recorded with a commercial EEG headband (Epoc+ from EMOTIV). The signals were acquired with a sampling frequency of 128 Hz. The channels that record the biosignal in the device used are based on the international 10–20 system; AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4.

The sampling of this database was carried out in a controlled environment and each subject participated in three sessions: one in total silence (silent), another with relaxing music (relaxing), and another with pleasant music chosen

by the subject participating in the session (pleasant). In each session, participants are asked to keep their eyes closed for 40 s, then open them and do basic multiplication exercises for 5 min. To induce a state of emotional stress, each exercise must be solved within a time limit of less than 5 s and if the answer is wrong, the participants get negative feedback. At the end of the mathematical test, the participants close their eyes for another 30 s to finish taking the sample.

This protocol for inducing stress is motivated by the interest of observing whether listening to music generates an effect on the perception of stress, such that it can be identified in brain waves, which can result in identifying stimuli that can be used in the management of stress and thus prevent its unwanted health effects. For details of this database, consult [7].

The signals were preprocessed for noise removal, to eliminate visible artifacts in the signal, such as thermal noise, electromagnetic interference, flickering and other muscular movements that influence the signal, following the steps described in Fig. 3.



**Fig. 3.** General steps for signal preprocessing.

The whole signal has a duration of almost 7 min, but the interest is focused in only one minute of the signal, from second 240 to second 300 as this is the segment when the participants are more concentrated on the mathematical task, the 14 channels recorded were selected for signal analysis. After useful information segment selection, the signal was subjected to the elimination of the average of the signal per channel, called baseline removal. Baseline removal eliminates electrical activity that is not related to the event being analyzed, can improve signal quality and reduce noise.

Then, the resulting signal was filtered with the intention of obtaining only the data within a range of 4–32 Hz, since the Theta, Alpha and Beta frequency bands are found in this interval. The patterns of electrical activity recorded in the brain are classified into five main frequency band types: Delta, Theta, Alpha, Beta and Gamma. Each type of brain wave has a characteristic frequency and amplitude and is associated with different mental and emotional states. In the study of stress, Alpha, Beta and Theta waves can be useful because they are associated with relaxation and meditation.

This filtering was followed by visual inspection and manual removal of certain types of noise by cutting out useless fragments of the signal. Independent

component analysis was applied to remove artifacts and channel interpolation was calculated to counter information removal. This process was carried out with the help of the EEGLAB MATLAB toolbox [16]. Although the removal of the baseline and the application of filters is done automatically, the cutting of useless segments of the signal is performed manually to prevent the cutting of relevant information. This process is secondary to the main objective of our work which is to propose the extraction of characteristics of the EEG signal by fuzzy granulation.

A common way to analyze EEG data is to decompose the signal into functionally distinct frequency bands, we propose fuzzy information granulation for feature extraction. First the signal was divided into granulation windows, each of these windows were used to extract granules *low*, *mid*, *high* as described above. The extracted granules are used as characteristics to represent samples of EEG signals used to classify three stress scenarios.

A classification model was generated for each participant, since stress is perceived by each person differently, and EEG signals present inter-subject variability. To classify the samples with 3 granules, exploratory experiments were carried out in which Support Vector Machine, Naive Bayes, Neural Network, Random Forest with 20, 50 and 100 trees classifiers were tested. Finally Random Forest (RF) with 50 trees was used because it was the classifier that achieved the best results. The accuracy percentages are obtained by 10-fold cross-validation, randomly partition the data into 10 sets of equal size, then using 90% of samples for training (9 partitions) and 10% (1 partition) for testing, then repeating this process 10 times.

### 3.1 Fuzzy Clustering

When using fuzzy clustering, 10s windows were segmented from the preprocessed unordered signal. 3 sets of samples were constructed: low granules, mid granules, high granules. Each sample has 14 features corresponding to the 14 original channels of EEG signal. Each set is composed of samples of the three classes: silent, pleasant and relaxing. Experiments were made classifying the three described stress scenarios generating separate models for each granule. Classification results are showed in Table 2.

Considering the model presented in [12] a model for each low, mid and high granule was constructed, and then the classification results were combined to emit a final classification for each sample. We trained different models for each set of granules with the same data. When we have new data, we will get a prediction of each model. Each model will have an associated vote. In this way, we will propose as a final prediction what most models vote. For Fuzzy Clustering granulation the results do not reach higher accuracy than when the granules are classified separately, as shown in Table 3.

To take advantage of the information from all the granules at the same time, a set of samples was created where each instance is characterized by 42 attributes, corresponding to the concatenation of 3 granules from each of the 14 channels

**Table 2.** Fuzzy C Means Granulation. Accuracy percentage achieved.

Subject	low	mid	high
1	98.11%	98.33%	93.33%
2	90.44%	93.33%	93.33%
3	97.66%	9766.00%	83.33%
4	93.33%	73.33%	96.67%
5	76.67%	60.00%	83.33%
6	97.33%	93.33%	96.67%
7	86.67%	98.67%	96.67%
8	97.66%	73.33%	76.67%
9	90.33%	73.33%	96.67%
10	80.00%	94.44%	86.67%
11	96.67%	88.89%	94.44%
12	86.67%	83.33%	86.67%
Average	90.96%	85.67%	90.37%

**Table 3.** Fuzzy C Means Granulation. Combination of classification results of the 3 granules. Classification accuracy percentage.

Subject	low/mid/high	Subject	low/mid/high
1	96.67%	7	93.33%
2	93.33%	8	73.33%
3	96.67%	9	96.67%
4	90.00%	10	83.33%
5	77.78%	11	94.44%
6	96.67%	12	88.89%
	Average	90.09%	

**Table 4.** Fuzzy C Means granulation. Granule concatenation. Accuracy percentage achieved.

Subject	low_mid_high	Subject	low_mid_high
1	98.66%	7	94.44%
2	93.33%	8	98.66%
3	93.33%	9	94.44%
4	96.67%	10	88.89%
5	93.33%	11	93.33%
6	98.66%	12	88.88%
	Average	94.39%	



of the EEG signal. With this group of samples the results shown in Table 4 were obtained.

The average classification accuracy obtained by concatenating granules is the highest for granulation with Fuzzy Clustering, reaching 94.39% compared to 90.09% achieved by combining the results of the models obtained for each granule set and 90.96% obtained with the low granule model.

### 3.2 Triangular, Trapezoidal and Minmax Granulation

To granulate the EEG signal based on MF, first the signal was segmented in 10 s granulation windows, then each granulation window was ordered in ascending order. Each of this windows where used to extract granules *low*, *mid*, *high* applying the MF described above. We worked with tree different data sets to model classifiers for each granule separately: the data set with granules low, the data set with granules mid and the data set with granules high. Each sample has 14 features corresponding to the 14 original channels of EEG signal. Classification results are showed in Table 5.

**Table 5.** Triangular, Trapezoidal and Minmax Granulation. Classification Accuracy Percentage.

	TG	TZG	mMG
low	91.66	86.67	89.89
mid	90.00	91.01	91.01
high	93.34	88.76	89.89

Subsequently, the classification results of each separate granule were taken and combined to give way to the results of the Table 6. In this case the classifications assigned by each classifier were taken as votes to assign the final classification of each sample. For instance, a sample is classified as silent class when using the model obtained with the low granules, while the same sample is classified as relaxing by the model obtained with the mid granules and as silence by the model generated by the high granules, therefore the assigned class is silent because this classification had two votes.

Also concatenated granules were used to classify three stress scenarios, each sample is described with 42 characteristics, corresponding to 3 granules of each of the 14 channels of the EEG signal, the classification accuracy results are shown in Table 7.

In [7] characteristics were extracted by calculating statistical measures (maximum, average, standard deviation, variance, statistical asymmetry and kurtosis) of the absolute power of the Alpha, Beta and Theta bands. Each of the bands was used separately, representing distinct sets of characteristic modeling three independent classifiers using Random Forest. Afterwards, the characteristics of

**Table 6.** Triangular, Trapezoidal and Minmax Granulation. Combination of Classification Results from low, mid and high Granules. Classification accuracy percentage.

	TG	TZG	mMG
low/mid/high	95.45	93.33	91.66

**Table 7.** Triangular, Trapezoidal and Minmax Granulation. Concatenation of low, mid and high Granules. Classification accuracy percentage.

	TG	TZG	mMG
low_mid_high	95.55	96.59	92.13

**Table 8.** The results reported in [7] focused on the same classification problem with the same database. Average percentage of classification accuracy.

	Alpha	Beta	Theta	Concatenated
Average Accuracy	79.5	92.2	78.9	93

the 3 bands were concatenated to model a single classifier also with random forest. The results obtained in the work of Reyes [7] are shown in Table 8:

As can be seen in Tables 7 and 8 the average accuracy classification obtained by concatenating the low, mid and high (low\_mid\_high) fuzzy granules reaches 96.59% while the concatenation of characteristics extracted from the Alpha, Beta and Theta bands have an average accuracy of 93%.

## 4 Conclusions

The present work proposes fuzzy granulation as a way to characterize EEG signals for the classification of different stress conditions. In total, we tested four different techniques of fuzzy granulation time series to extract features from EEG signals: Fuzzy C-Means, Triangular Mf, Trapezoidal Mf and Minmax granulation. Granulation of time series to extract features from EEG signal is an approach that allows finding relevant information in electroencephalographic signals recorded in different stress scenarios. These feature extraction proposal leads to encouraging results in multi-class stress classification. Accuracy percentage classification of 96.59% has been achieved by extracting three granules from the EEG signal using TZG, generating a data set where each EEG sample is characterized by the concatenation of the extracted granules, as shown in Table 7.

The study of the effects of music on stress has long been of great interest, for example in [17] it was found that preoperative music can help normalize hypertensive responses during outpatient surgical stress, in [18] data show that in the presence of music, the level of salivary cortisol stopped increasing after psychological stress, in [19] is presented a mobile application that used yoga therapy and music to help users relax. However none of these works have the

focus of using machine learning or fuzzy granulation for automatic classification of different stress patterns.

By being able to clearly differentiate the three scenarios in which the EEG signal sampling was done, total silence, relaxing music and pleasant music; it is possible to assure that listening to music has effects on the body's reaction to a cognitive challenge. This allows us to consider music as a tool with the potential to reduce the negative effects of stress in its early stages and thus mitigate its negative effects on health. More research is needed in this regard in order to effectively identify stimuli that reduce stress.

Regarding the results obtained with Fuzzy Clustering, with an average classification accuracy that is within the range of 85.67% to 94.39%, it can be said that they are below the results achieved with the MF granulation approach. Changing multi-class approach to a binary classification could improve the results. This is because binary classification algorithms are simpler and easier to implement than multi-class classification algorithms. In addition, binary classification algorithms may be more accurate than multi-class classification algorithms when classes are not well separated, there is less chance of error and less complexity in the model. Furthermore testing the fuzzy granulation on the preprocessed signal, in smaller time windows could also increase accuracy classification results. All these experiments are proposed as future work.

## References

1. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and Fuzzy Logic. *Fuzzy Sets Syst.* **90**(2), 111–127 (1997)
2. Ouyang, T., Pedrycz, W., Reyes-Galaviz, O.F., Pizzi, N.J.: Granular description of data structures: a two-phase design. *IEEE Trans. Cybern.* **51**(4), 1902–1912 (2021)
3. Huang, Z., Li, J.: Feature subset selection with multi-scale fuzzy granulation. *IEEE Trans. Artif. Intell.* **1** (2022)
4. Lu, W., Pedrycz, W., Liu, X., Yang, J., Li, P.: The modeling of time series based on fuzzy information granules. *Expert Syst. Appl.* **41**(8), 3799–3808 (2014)
5. Subhani, A.R., Mumtaz, W., Saad, B.M., Kamel, N., Malik, A.S.: Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access* **5**, 13545–13556 (2017)
6. Zanetti, M., et al.: Multilevel assessment of mental stress via network physiology paradigm using consumer wearable devices. *J. Ambient. Intell. Humaniz. Comput.* **12**(4), 4409–4418 (2019)
7. Reyes Galaviz, R.S.: Análisis con electroencefalografía (EEG) de la escucha de música para el estudio de estrés académico. Item 1009/2048: Repositorio INAOE (2021)
8. Gu, X., et al.: EEG-based brain-computer interfaces (BCIs): a survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **18**(5), 1645–1666 (2021)
9. AlShorman, O., et al.: Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection. *J. Integr. Neurosci.* **21**(1), 020 (2022)

10. Tsai, Y.-H., Wu, S.-K., Yu, S.-S., Tsai, M.-H.: Analyzing brain waves of table tennis players with machine learning for stress classification. *Appl. Sci.* **12**(16), 8052 (2022)
11. Liu, L., Ji, Y., Gao, Y., Li, T., Xu, W.: A novel stress state assessment method for college students based on EEG. *Comput. Intell. Neurosci.* **2022**, 1–11 (2022)
12. Zhang, R., Shen, F., Zhao, J.: A model with fuzzy granulation and deep belief networks for exchange rate forecasting. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 366–373 (2014)
13. Ruan, J., Wang, X., Shi, Y.: Developing fast predictors for large-scale time series using fuzzy granular support vector machines. *Appl. Soft Comput.* **13**(9), 3981–4000 (2013)
14. Reyes-Galaviz, O.F., Pedrycz, W.: Granular fuzzy models: analysis, design, and evaluation. *Int. J. Approximate Reasoning* **64**, 1–19 (2015)
15. Pedrycz, W., Homenda, W.: Building the fundamentals of granular computing: a principle of justifiable granularity. *Appl. Soft Comput.* **13**, 4209–4218 (2013)
16. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**(1), 9–21 (2004)
17. Allen, K.J., et al.: Normalization of hypertensive responses during ambulatory surgical stress by perioperative music. *Am. J. Hypertens.* **11**(4), 19A (1998)
18. Khalfa, S., Bella, S.D., Roy, M., Peretz, I., Lupien, S.J.: Effects of relaxing music on salivary cortisol level after psychological stress. *Ann. N. Y. Acad. Sci.* **999**(1), 374–376 (2003)
19. Vijayaragavan, G.R., Raghav, R.L., Phani, K.P., Vaidyanathan, V.: EEG monitored mind de-stressing smart phone application using yoga and music therapy. In: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Greater Noida, India, pp. 412–415 (2015)