Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Multistage Anxiety State Recognition based on EEG Signal using Safe-Level SMOTE

To cite this article: Tee Wee Shing *et al* 2023 *J. Phys.: Conf. Ser.* **2622** 012010

Multistage Anxiety State Recognition based on EEG Signal using Safe-Level SMOTE

Tee Wee Shing[1], Rubita Sudirman[1], Syarifah Noor Syakiylla Sayed Daud[1*], Mohd Azhar Abdul Razak[1], Nor Aini Zakaria[1], Nasrul Humaimi Mahmood[1]

[1]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

*Email: sya.syakiylla@gmail.com

Abstract. Anxiety is a complicated emotional condition that has a detrimental effect on people's physical and mental health. It is critical to accurately recognize anxiety levels in early stage. The anxiety can be detected by pattern of brain signal using brain imaging tools. However, the common problem with dataset acquired from brain is imbalanced class distribution. Hence, the purpose of this work is to mitigate the imbalanced class distribution issue by removing data outlier and using improved Synthetic Minority Oversampling Technique (SMOTE) for improving the classification performance. This work used of the freely accessible Database for Anxious States based on Psychological stimulation (DASPS) that comprises of 14 channels electroencephalography (EEG) signal. It acquired from 23 subjects when they were exposed to psychological stimuli that elicited fear. The DASPS need to be processed for removing noises, extracting important features and sampling with Safe-level SMOTE method. Then, the processed DASPS was categorized into three types of model: Model A, Model B, and Model C. The feature Model C from enhanced DASPS class distribution obtained the precision of 89.7% and accuracy of 89.5% using optimized *k*-nearest neighbour (*k*-NN) algorithm. The proposed method showed outstanding classification performance than others existing methods in recognizing multistage anxiety.

1. Introduction

Currently, clinical symptoms or a set of questionnaires, such as the Beck Anxiety Inventory, the Self-Rating Anxiety Scale, the State-Trait Anxiety Inventory, the Hamilton Anxiety Scale, and the Manifest Anxiety Scale, are used to make the diagnosis of anxiety disorder. However, this diagnosis approach is not objective enough and is vulnerable to the possibility of subjective bias [1]. This might degrades the overall quality of the evaluation. However, many researchers have found that the subjective feelings of anxiety, stress, and emotion can be quantified objectively using a variety of techniques to acquire physiological signals from individuals. Among of the techniques are electroencephalography (EEG), electrocorticography, and functional magnetic resonance imaging (fMRI). The main reason of this technique used for research is because the physiological signals are harder to be control or manipulated by individual when facing stress situation. The use of EEG based technologies to monitor the individual mental health and cognitive performance is becoming more widespread, and researchers have investigated many approaches to diagnosing the anxiety disorder. The DASPS was proposed by Baghdadi et al., [2] where it contains EEG information from 23 subjects during anxiety elicitation in six different settings. The time, frequency and time-frequency domain features were extracted from DASPS and feed into the Stacked Sparse Autoencoder for classification purpose. They had found that the anxiety was well elicited in 1 second with the maximum accuracy for 2-level of 83.5% and 4-level of 74.6% classes of anxiety based *k*-NN and support vector machine (SVM) classifier. In another work, Shikha et al., [3] applied recursive feature elimination with cross-validation (RFECV) with the classifiers on the publicly available DASPS database to decrease redundancy between features and improve results. They found that the classification accuracy was 83.9% for 2-level class of anxiety. Their result showed enhancement of around 5% for time-domain, 8% for time-frequency domain, and 0.5% for all features combination. Table 1 summarized the other related studies of anxiety and stress classification using EEG signal.

Table 1. Past studies related to anxiety and stress classification using EEG signal

EEG databases	Methodology	Results	References
A database for emotion analysis using physiological signals (DEAP)	Genetic algorithm (GA) with *k*-NN	71.8% of accuracy in recognize stress and calm	Hasan et al., [4]
DEAP	Boruta with *k*-NN	73.4% of accuracy in recognize stress and calm	Jang et al., [5]
DEAP	Borderline-SMOTE with Convolutional Neural Network (CNN)	97.5% of accuracy for three stages emotional recognition	Chen et al., [6]
Conduct public English speech to simulate anxiety	Support Vector Machine (SVM)	62.6% accuracy for four class of stress level	Li et al., [7]

However, these solutions above, to some part, overlook a crucial issue related to the treatment of data imbalance. The imbalance dataset quantity of samples in different classes of the original dataset may vary substantially, especially in the multi-classification process. It will have a direct impact on the final model performance [8]. Therefore, this project will focus on using computational approach based supervised machine learning to create a model to classify the anxiety state level of an individual using EEG signals. The Safe-level SMOTE was applied in the pipeline of machine learning to tackle the imbalanced DASPS class distribution issue.

2. Methodology

2.1. EEG dataset description

This project employs the publicly accessible DASPS dataset for the purpose of detecting anxiety level, which was gathered by Baghdadi et al., [2]. This database consists of raw EEG data (.mat format) acquired from the 23 participants. Ten of them were male and 13 of them were female with average age of 30 years old. The Emotiv EPOC wireless EEG headset based on 14 channels and two mastoids was used to measure and record the brain signal of participant during mental assessment. The frequency sampling rate of acquired EEG signals was 128 hertz. The study stimulus anxiety was using in-vivo exposure therapy. The assessment was divided into six different situations and each situation was divided into two parts. The first was the psychotherapist recites the situation for the first 15 seconds. Then, the next part was the subject recalls it for the next 15 seconds. After each situation, the participants need to rate their felt during assessment based on Self-Assessment Manikin (SAM). The participants were instructed to closed eyelids and keep minimal motions while the EEG data was being recorded.

2.2. Pre-processing of raw EEG signal
To begin, the extrinsic signal artifacts were removed using a FIR bandpass filter with a cut-off frequency of 45 Hz and a cut-off frequency of 4 Hz to eliminate the high-frequencies and low-frequencies artifacts. These artifacts usually occurred due to power line interference, excessive motion, and eyes blinks and movements. It is crucial to identify artifacts components in order to minimize inherent signal artifacts. Therefore, the raw EEG signals were decomposed using the Informax independent component analysis (ICA) approach. The IClabel EEGLAB plugin classified the components into six categories which are brain, muscle, eyes, heart, line noise, channel noise, and others. Subsequently, those components containing most non-brain components will be removed. Finally, the automatic artifact removal in the toolbox is applied to the processed signal for improving the quality of data. This technique removed bad data sections, which unable to be eliminated via both ICA and FIR filtered. The entire procedure was automated by an EEGLAB script.

2.3. Features extraction
The main intend of extracting features are to reduce the losing of important attributes consisted within the interest signal. It converts the signal into the identifiable measurement, a distinguishing attribute, and a functioning component by reducing the number of resources from large dataset. This was necessary to reduce the implementation complexity, cut-off the cost for signal processing, and to prevent the necessity for data compression. In this work, the hybrid pools of the features is from three distinct domains which are time domain (Higuchi's Fractal Dimension and Hjorth parameter: complexity, mobility, and activity), frequency domain (spectral entropy and sample entropy), and time-frequency domain (absolute power of frequency bands, alpha/beta ratio, theta/alpha ratio, power spectral density of frequency bands, and root mean square of frequency bands). The discrete wavelet transform (DWT) is used in the time-frequency domain to extract both the time and frequency components of the data. Wavelet transforms provide a higher level of resolution than Fourier Transforms. The DWT based Daubechies function of order 4 with five decomposition level was used extract the EEG rhythms from DASPS dataset. This wavelet attributes was chosen because it well suited for processing biomedical signals [9]. The input signal will pass through low-pass and high-pass filter, which divided the frequencies into two bands of low-pass component (approximation component) and high-pass component (detail component). Figure 1 showed the DWT decomposition tree for extracting EEG signal with 128 Hz of frequency sampling. After decomposition, the EEG rhythms absolute power were obtained using Welch's approach through estimation of power spectral density. Each decomposed signal also contains a statistical feature called the Root Mean Square.

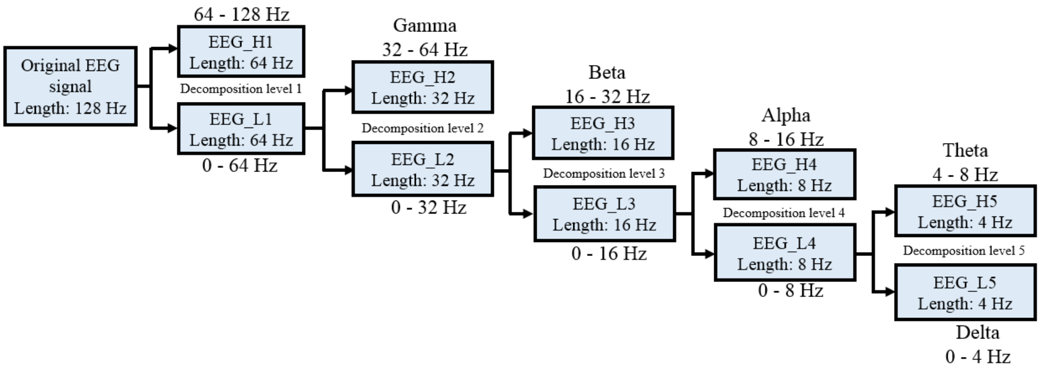

Figure 1. Decomposition tree of discrete wavelet transform for extracting time and frequency features of EEG rhythms [10]

2.4. Data augmentation
The data augmentation is an artificial process to increase the number of samples and datapoints by yielding a new samples and datapoints from existing data. In this work, the safe-level SMOTE method

was used for data augmentation to compensate the inequitable class distribution of DASPS dataset. This step is critical because the minority classes of light and moderate are 5 to 10 times lesser compared to majority classes of severe and normal anxiety state. In addition, a few classes of DASPS dataset are overlap and positioned next to categorization boundary. In this regard, it caused low classification performance. The safe-level SMOTE was operating based on *k*-NN rules and selection of appropriate properties is necessary to acquire promising augmentation output. Three main steps taken using *k*-NN rules which are calculating the distance between neighbors, identify the *k* closest neighbors to determine the bias-variance trade-off related with solve of underfitting and overfitting issue, and finally select the new label of samples. The Bayesian optimization will chose the optimal hyperparameter for the *k*-NN model. It had been found that the most optimal *k*-value was 5. The basic steps implement for augmented DASPS using safe-level SMOTE method were locate the criteria area to yield the synthetic data and produce the synthetic dataset based on a safe-level ratio. It can be summarized as below:

1. Calculate the distance of nearest neighbors (n) and minority class instances in observe set (p) using Euclidean distance.
2. Executing the safe-level SMOTE method based on k-value of 5.
3. Chose one of k nearest neighbors generated from the minority class.
4. Recalculate the distance between n and its neighbors with the same k using Euclidean distance.
5. Chose one of the k nearest neighbors generated from the minority class.
6. Determine the safe-level to p, n, sl_p, and sl_n as represents in equations (1) and (2).
   sl_p = the number of positive instances in k – nearest neighbors for p (1)
   sl_n = the number of positive instances in k – nearest neighbors for n (2)
7. Determine the safe-level ratio for p and n as in equation (3) and categorized into respective cases as in Table 2. Then, the new synthetic instance (si) was generated in the coordinate of *si* as shown in equation (4).
   $$safelevel\_ratio, sl\_ratio = sl_p/sl_n \quad (3)$$
   $$si = pi+(gap\times diff) \quad (4)$$
8. Calculate the difference between p and n.
9. Refer to the range of random numbers based on the safe level ratio (*sl_ratio*) in Table 2.
10. The difference acquired in step 8 was multiplied by a random number in step 9.
11. The result yield from step 10 was adding to p for generating the new instances.
12. The steps were repeated until the number of minority class observations was approximately similar to that of majority class observations.

2.5. Data classification and evaluation
Several methods were used for classified and evaluated before and after DASPS dataset augmented. Detail explanation as discussed below.

2.5.1. k-NN algorithms
The *k*-NN technique is then used to classify the pool of extracted features. This has the advantage of simplifying model and minimizing complexity of process for small datasets. It is efficient to utilize for a small, featured dataset because of its non-parametric properties and samples classification was based on the votes of the *k*-nearest neighbors [11]. After each classification, Bayesian optimization is implemented to select the maximized hyperparameters for the *k*-NN model. This optimization minimized the scalar objective function f(x) for x in a bounded domain. The x represents the different value of hyperparameters of *k*-NN which includes the number of nearest neighbors, standardization, distance, weight, and exponent [12].

Table 2. Rule of synthesis minority instance in safe-level SMOTE [10]

Cases	sl_ratio	slp	Synthesis at a range between p and n, gap	Description
1	=∞	0	do not produce a positive synthetic instance	Both p and n instances are noise. Thus, no synthetic data are generated.
		≠0	gap = 0	The n instance is noise; therefore, synthesis is as close as possible to the location of the p instance, and the synthetic data will be generated far from n by duplicating p.
2	=1		0 ≤ gap ≤ 1	The safe level of p instance is the same as with n instance. The synthetic data will be generated along the line between p and n because p is as safe as n.
3	>1		0 ≤ gap ≤ 1/sl_ratio	The p instance is safer than the n instances; therefore, synthesis is closer to the p position. The synthetic data will be generated closer to p at a distance [0,1/SLR].
4	<1		(1 – sl_ratio) ≤ gap ≤ 1	The n instance is safer than the p instance; therefore, synthesis is closer to the n position. The synthetic data will be generated closer to n at a distance [1-SLR,1].

2.5.2. k-fold cross validation

Finally, the performance of the DASPS features model is determined based on the k-fold cross validation. This work used a 10-fold cross validation, where the dataset was randomly categorized into 10-fold. Each of fold acquires 1 over 10 of the datasets for validation set and the remain dataset was for training set. The validation set was shifting for each iteration of fold. The remain data was feeding into the classification model for training. The final accuracy of model performance was obtained from average accuracy result from each iteration fold [13].

2.5.3.Model performance evaluation metrics

The performance of classifier in this work was determine via a confusion matrix. There are four classification performance measures in confusion matrix which are accuracy, precision, F-value, and true positive rate (TPR). The accuracy refers to the number of correct predictions yielded by model via the entire test dataset, whereas the precision is an accuracy of a positive prediction. Besides, the precision also indicates the reliability of a predicted positive result. It can be controlled through tuning of model hyperparameters and parameters. Meanwhile, the F-value is a metric for determining the correctness of a model based on its precision and recall. The TPR is a metric that determines the number of expected true positive (TP) from all positives in dataset instead the expected number of false positives (FP). The classification parameter used in this work was represented in equation (5)-(8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1\ score = \frac{Precision \bullet TPR}{Precision+TPR} \tag{8}$$

Three models obtained from this work which are Model A (dataset from original DASPS dataset), Model B (synthetic dataset from Model A after augmented using safe-level SMOTE method), and Model C (synthetic dataset from Model C after outlier removal). The outlier removal method refers to the process of removing bad data point that far from average value of others. The z-score was performed to remove the bad data point. Then, the classification results from these models were compared and evaluated.

3. Results and Discussions

This section was discussed on outcomes obtained from classification performance of DASPS dataset models.

3.1. Classification performance of Model A

Table 3 represents the classification performance from different domain features of DASPS dataset. The main purpose was to determine and select the optimal classification performance of domain features by *k*-NN to be further processed with safe-level SMOTE method and outlier removal.

Table 3. The classification performance of different features from DASPS dataset before augmented and outlier removal

Features	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Time	63.2	58.1	41.7	43.8
Frequency	60.0	29.6	32.1	30.0
Time-frequency	**63.2**	**55.9**	**52.7**	**54.1**
All	61.6	45.9	37.3	38.1

It shown that the frequency domain obtains the overall lowest performance with only 60.0% and 30.0% of accuracy and F1-score respectively. It only manages to recognize the majority classes which is ‘normal and ‘severe’. The model with all feature domain gained the lower classification performance than expected. This model achieved accuracy of 61.6% and F1-score of 38.1%, which was slightly lowered than the time-frequency domain features. Meanwhile, the time and time-frequency features achieved similar accuracy of 63%, but yielded different value for other classification parameter. The time-frequency features having 10% higher of F1-score (54.1%) compared to time-domain features (43.8%). Therefore, it suggested that the time-frequency domain was better than other features domain and this features was selected to be augmented with safe-level SMOTE and undergo outlier removal process. Through features selection, more time can be saved and less architecture complexity instead of performed data improvement on each features.

3.2. Classification performance of Model B

Figure 2 showed the classification performance of time-frequency domain features after augmented using safe-level SMOTE. The classes distribution was improved and balanced especially for minority classes of light and moderate anxiety state. It was 80 and 94 new samples had been successfully synthesized for the light and moderate anxiety states, respectively. Hence, the number of samples was more evenly distributed in the DASPS dataset. The percentage improvement of class distribution for each states was: normal (52% to 31%), light (8% to 24%), moderate (4% to 21%), and severe (36% to 24%).

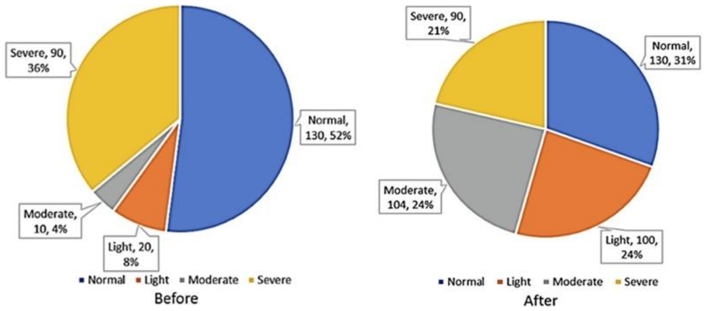

Figure 2. Class distribution before and after augmented using safe-level SMOTE method

Table 4 indicated the classification performance for Model B. This model was generated from Model A that augmented using safe-level SMOTE method. It had significantly improved performance than Model A of about 16.3%. The others parameter also showed improvement such as precision (23.1%), recall (27.3%), and F1-score (41.5%) in comparison to Model A. Therefore, it showed that the Model B obtained better and balanced classes distribution, which revealed the safe-level SMOTE was efficient for augmented DASPS dataset. The class imbalanced issue had been tackle by using this method. Then, the effect of adding an outlier removal method of rejecting bad data point on classification will be known in the following section.

Table 4. Classification performance of Model B

Model accuracy: 80% (+16.3)			
Anxiety states	Precision (%)	Recall (%)	F1-score (%)
Normal	71.8	65.6	68.6
Light	89.8	97.0	93.3
Moderate	94.3	96.2	95.2
Severe	60.2	62.2	61.2
Overall	79.0 (+23.1)	80.0 (+27.3)	79.6 (+41.5)

3.3 Classification performance of Model C

The classification performance of Model C aims to determine the effect of implemented safe-level SMOTE method and outlier removal. Based on Table 5, it can be stated that the classification performance of Model C was better than Model B. The accuracy of Model C was improved about 10% than Model B. All of the indicators also showed better result, where the percentage increment of precision was 10.7%, recall was 9.0%, and F1-score was 9.8%. Therefore, it indicated that combination of those method yield better classification performance, which considered as a good model since it has high F1-score and accuracy.

Table 5. Classification performance of Model C

Model accuracy: 89.5% (+10.0)			
Anxiety states	Precision (%)	Recall (%)	F1-score (%)
Normal	92.3	75.8	83.2
Light	86.7	96.6	91.4
Moderate	91.7	94.6	93.1
Severe	87.9	92.0	89.9
Overall	89.7 (+10.7)	89.0 (+9.0)	89.4 (+9.8)

4. Conclusion

The recent research showed an enhanced classification performance for different states of anxiety level from DASPS dataset. A safe-level SMOTE method was used to handle the imbalanced classes distribution in dataset. In addition, the outlier removal based z-score method also employed to improved classification performance. A few steps were used to process the DASPS dataset which are acquire raw DASPS dataset, artefact elimination, features extraction, data augmentation, features classification and model evaluation. The result of the modified dataset achieves 89.7% and 89.4% of accuracy and F1-score respectively by *k*-NN classifier on the four-class anxiety classification. In conclusion, the proposed method improved the accuracy to 15.1% compared to the existing method. Thus, this proposed methodology can be implemented to identify anxiety states and other imbalanced dataset.

Acknowledgments

The authors would like to thank the Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia for their support under FRGS grant R.J130000.7851.5F425, UTMER grant Q.J130000.3851.20J75 and PDRU grant Q.J130000.21A2.05E52.

References

[1] Zheng Y, Wong T, Leung B and Poon C 2016 Unobtrusive and multimodal wearable sensing to quantify anxiety *IEEE Sens. J.* **16** 3689–96.

[2] Baghdadi A, Aribi Y, Fourati R, Halouani N, Siarry P and Alimi A 2021 Psychological stimulation for anxious states detection based on EEG-related features *J. Ambient Intell. Humaniz. Comput.* **12** 8519–33.

[3] Shikha S, Agrawal M, Anwar M and Sethia D 2021 Stacked sparse autoencoder and machine learning based anxiety classification using EEG signals *Int. Conf. AI-ML Sys.* **7** 1–7.

[4] Hasan M and Kim J 2019 A hybrid feature pool based emotional stress state detection algorithm using EEG signals *Brain Sci.* **9** 376–91.

[5] Shon D, Im K, Park J, Lim D, Jang B and Kim J 2018 Emotional stress state detection using genetic algorithm-based feature selection on EEG signals *Int. J. Environ. Res. Public Health* **15** 2461–72.

[6] Chen Y, Chang R and Guo J 2021 Effects of data augmentation method bordeline SMOTE on emotion recognition of EEG signals based on convolutional neural network *IEEE Access* **9** 1–9.

[7] Li Z, Wu X, Xu X, Wang H, Guo Z, Zhan Z and Yao L 2022 The recognition of multiple anxiety levels based on electroencephalograph *IEEE Trans. Affect. Comput.* **13** 519–29.

[8] Par O, Akcapinar S and Sever H 2019 Small and unbalanced data set problem in classification *27th Signal Process. Comm. App. Conf.* 1–4.

[9] Cheong L, Sudirman R and Hussin S 2015 Feature extraction of EEG signal using wavelet transform for autism classification *ARPN J. Eng. Appl. Sci.* **10** 8533–40.

[10] Daud S, Sudirman R and Shing T 2023 Safe-level SMOTE method for handling the class imbalanced problem in electroencephalography dataset of adult anxious state *Biomed. Signal Process. Control* **83** 1–12.

[11] Uddin S, Haque I, Lu H, Moni M and Gide E 2022 Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction *Scientific Reports* **12** 1–11.

[12] Greenhill S, Rana S, Gupta S, Vellanki P and Venkatesh S 2020 Bayesian optimization for adaptive experimental design: A review *IEEE Access* **8** 13937–48.

[13] Berrar D 2018 Cross-validation *Ency. Bioinform. Comput. Biol: ABC of Bioinform.* 542–45.