International Journal of Pioneering Technology and Engineering

Research Article

Application of Dimension Reduction Methods for Stress Detection

Erhan Bergil [a*]

[a] Faculty of Engineering, Department of Electrical Electronics Engineering, Amasya University, Amasya, Türkiye

Article Info

Article history
Received: 31/10/2023
Revised: 18/11/2023
Accepted: 05/12/2023

Keywords:
Feature Selection,
Dimension Reduction,
Stress Detection.

ABSTRACT

Effective detection of stress situations plays an important role in combating it. This is the main source of motivation for research to identify and evaluate different psychological conditions. Different monitor signals are used to identify individuals' stress situations in daily life. Electroencephalogram (EEG) signals are the main component used to detect stress and depression. The long-term acquisition of this signals partially interrupts daily life and negatively affects it. Researchers are trying to develop wearable technologies that can eliminate this disadvantage. In this study, stress situations are detected utilizing different sensors without EEG signals. The achievements of three different classification methods for different dimensional feature spaces have been compared. The effects of the feature selection and dimension reduction methods on the system performance have been analyzed. During the dimension reduction process, Minimum Redundancy Maximum Relevance (MRMR), Anova, Chi-2, Relieff, Kruskal Wallis (KW) and Principal Component Analysis (PCA) methods are implemented. Support Vector Machines (SVM), Linear Discriminant Analysis (LDA) and k-Nearest Neighbor (k-NN) methods are used as classifier. The best performance is achieved with 96.2 % accuracy in 15-dimensional by using LDA and PCA methods together.

1. Introduction

Anxiety is a feeling of worry, fear, or distress felt to a mild, moderate, or severe degree in response to perceived or impending stress or threat factors. Stress may be prevented or lessened from having detrimental impacts on one's health, thus it's critical to recognize both the mental and physical reactions to stress in order to assess and manage it. Research is carried out to monitor neurological status and detect stress situations. The main goal is to combat anxiety by using wearable technologies and designing systems that will not affect the daily lives of individuals. Living organisms produce their own monitor signals. Using these signals, it is possible to monitor tissues, organs, or systems. Electroencephalogram (EEG) signals are the main component used in neurological status detection. Long-term EEG recording from a wearable device, it is unsuitable and uncomfortable during daily activities. At the same time, it has a certain cost. For these reasons, it is important to evaluate the neurological status using different sensors without EEG [1-3].

Priya et al. realized prediction of anxiety, depression and stress utilizing machine learning. While the best performance was achieved with random forest, it was seen that the naive bayes method came to the fore in terms of classification accuracy [4].

Pham detected neurological status 98.75% with accuracy utilizing tensor decomposition and machine learning. Linear Discriminant Analysis (LDA) classifier was more successful than the Multinomial Logit Regression method [5]. Rastgoo et al. used multimodal deep learning to classify driver stress levels. The effect of window size on system performance was shown [6].

Arpaia et al. performed real time stress analysis utilizing a wearable EEG instrument. As a result of the study using Support Vector Machines (SVM), k-Nearest Neighbor (k-NN), Random Forest (RF) and Artificial Neural Networks (ANN) classifiers, reached more than 90% accuracy [7]. Bobade et al. detected stress using machine learning and deep learning methods. In the study, acceleration (3-axes), electrocardiogram, blood volume pulse, body temperature, respiration, electromyogram and electrodermal activity measurements were used for feature extraction. An accuracy of up to 95.21% was achieved with the classification processes performed in 2-class and 3-class [8].

Jaloli et al. implemented neurological status classification using a convolutional neural network. In the study where the classification performance of the convolutional neural network method was compared with the performance of traditional methods such as support vector machines and random forest, a classification accuracy of 97.46% was achieved [9]. Masood et al. proposed a method based on fully convolutional short-term memory network (FCN-LSTM) to detect neurological status. Using this method and conventional methods, non-invasive sensor data were classified into 2-class and 4-class. The method performed with an accuracy of 98.6% for 4-class [10]. Iqbal et al. performed unsupervised and supervised classification for physiological stress detection. Accuracies of up to 75.0% were obtained for the methods tested on two data sets [11].

In this study, unlike previous studies, the effects of feature space on stress detection were analyzed. Feature spaces of different dimensions were obtained by using different feature selection and dimension reduction methods. At this point, techniques such as Principal Component Analysis (PCA), Anova, Chi-2, Relieff, Kruskal Wallis (KW), and Minimum Redundancy

***Corresponding author:** Erhan Bergil
*E-mail address: erhan.bergil@amasya.edu.tr
https://doi.org/10.56158/jpte.2023.56.2.02

Maximum Relevance (MRMR) were applied. Classification was carried out using LDA, SVM and k-NN methods. Thus, the performances of feature selection and dimension reduction methods were compared. A similar comparison was made for classification methods.

2. Material and Methods

2.1. Data Set

The data used in the study were obtained from the open-access physionet database [12]. This database contains non-EEG physiological signals that were gathered at the University of Texas at Dallas' Quality of Life Laboratory and used to infer the neurological status of 20 healthy volunteers, including their levels of physical stress, cognitive stress, emotional stress, and relaxation. The information includes electrodermal activity (EDA), temperature, acceleration (for x, y and z axes), heart rate (HR), and arterial oxygen saturation (SpO2) and was gathered using non-invasive wrist-worn biosensors. The sampling frequency is 1 Hz for HR and SpO2. The other measurements are sampled at 8 Hz. The physical stress state is achieved by standing and walking in the running band at different speeds. For cognitive stress, a reverse seven extraction and stroop test was used. A clip from the Zombie Apocalypse movie is shown for emotional stress. Figure 1 displays the measurements for a subject. Information about the participants is given in Table 1.

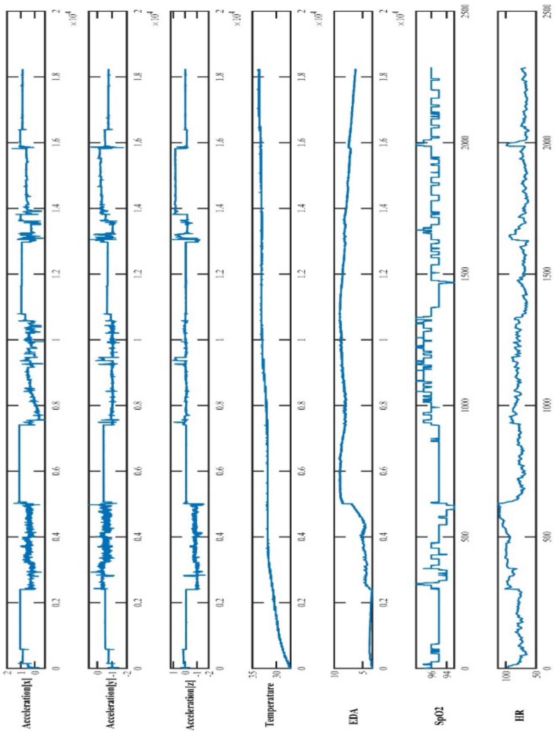

Fig. 1. Sensor outputs

Table 1. Participant information

Subject	Age	Gender	Height	Weight
1	30	M	177	94
2	28	M	172	68
3	28	M	177	91
4	22	M	167	58
5	30	M	182	82
6	30	F	167	58
7	33	F	157	90
8	27	M	182	64
9	25	M	177	68
10	23	M	180	64
11	26	M	170	71
12	32	F	162	53
13	29	F	167	64
14	19	F	160	50
15	23	M	165	64
16	24	M	180	54
17	23	M	167	57
18	23	M	177	64
19	22	M	167	64
20	24	F	160	44

2.2. System Flowchart

Analysis consists of preprocessing, feature extraction, dimension reduction and classification stages. The block diagram of the system is shown in Figure 2.

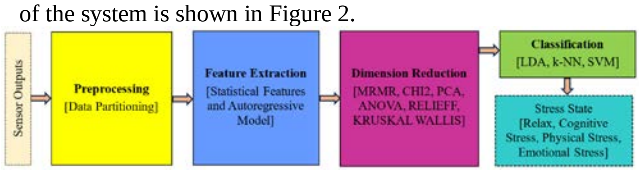

Fig.2. Block diagram of designed system

In the preprocessing stage, the data from the sensors are divided into relax state, emotional stress, cognitive stress and physical stress according to the conditions in which they were obtained.

2.2.1 Feature Extraction

Statistical features are obtained by calculating the maximum, minimum, mode and standard deviation of the measurements of the sensor outputs. Thus, a total of 28 statistical features are generated for 7 sensors. In addition to these features, 8th order autoregressive model coefficients are calculated for Acc[x] and SpO2 utilizing the Burg method [13-14]. The total number of features has been reached 44, including 16 features based on the autoregressive model.

2.2.2. Dimension Reduction

The 44-dimensional feature set is transformed into spaces with the dimensions of 5 and its multiples using different methods.

In this step, MRMR, CHI2, RELIEFF, KW feature selection algorithms and PCA methods, which are frequently used in the literature, are implemented [15-19]. Feature sets obtained in different dimensions are given as input to the classifiers.

2.2.3. Classification

Feature sets of different dimensions are classified using the 10-fold cross-validation method. Commonly used LDA, SVM and k-NN methods are applied as classifiers. [20-22]. As a result of these processes, relaxation, physical stress, cognitive stress and emotional stress are detected. The performances of classifiers and dimensionality reduction methods have been evaluated together.

3. Experimental Results

Feature spaces created in different dimensions are classified using the 10-fold cross validation method. The accuracies of classification procedures for implemented methods are given in Table 2.

Table 2. The performances of classifiers and feature selection methods

	MRMR			CHI2			RELIEFF		
Space Dimension	LDA	SVM	k-NN	LDA	SVM	k-NN	LDA	SVM	k-NN
5	87.5	91.2	85.0	81.2	88.8	83.8	83.8	86.2	81.2
10	87.5	87.5	72.5	90.0	90.0	85.0	86.2	88.8	80.0
15	91.2	87.5	71.2	92.5	88.8	86.2	83.8	80.0	73.8
20	87.5	90.0	78.8	91.2	88.8	80.0	85.0	81.2	71.2
25	86.2	92.5	76.2	88.8	85.0	82.5	87.5	76.2	68.8
30	83.8	87.5	78.8	90.0	86.2	78.8	86.2	80.0	75.0
35	87.5	82.5	72.5	88.8	87.5	78.8	82.5	81.2	80.0
40	88.8	88.8	75.0	87.5	85.0	77.5	86.2	81.2	80.0
mean	87.5	88.4	76.3	88.8	87.5	81.6	85.2	81.9	76.3
	ANOVA			KRUSKAL WALLIS			PCA		
Space Dimension	LDA	SVM	k-NN	LDA	SVM	k-NN	LDA	SVM	k-NN
5	90.0	87.5	92.5	85.0	86.2	83.8	70.0	68.8	62.5
10	88.8	87.5	87.5	92.5	90.0	88.8	76.2	71.2	61.3
15	92.5	88.8	85.0	91.2	91.2	86.2	**96.2**	81.2	61.3
20	87.5	91.2	82.5	90.0	90.0	81.2	88.8	85.0	70.0
25	85.0	90.0	78.8	87.5	88.8	78.8	85.0	82.5	63.7
30	88.8	87.5	81.2	86.2	90.0	80.0	85.0	80.0	63.7
35	88.8	88.8	81.2	88.8	88.8	80.0	90.0	73.8	57.5
40	86.2	81.2	80.0	85.0	82.5	81.2	86.2	70.0	**56.2**
mean	88.5	87.8	83.6	88.3	88.4	82.5	84.7	76.6	62.0

When Table 2 is examined, LDA is the most successful classification method according to the average classification accuracy. With this classifier, the highest average accuracy results are achieved in all dimension reduction methods except MRMR and KW. LDA classifier is followed by SVM and k-NN methods, respectively. The performance ranking of dimension reduction methods according to the average classification accuracies of the three classifiers was ANOVA (86.6%), KW (86.4%), CHI2 (86.0%), MRMR (84.1%), RELIEFF (81.1%) and PCA (74.4%). To compare classifier performances in different dimension spaces, the bar graphs obtained from the classification accuracies are shown in Figure 3-5.

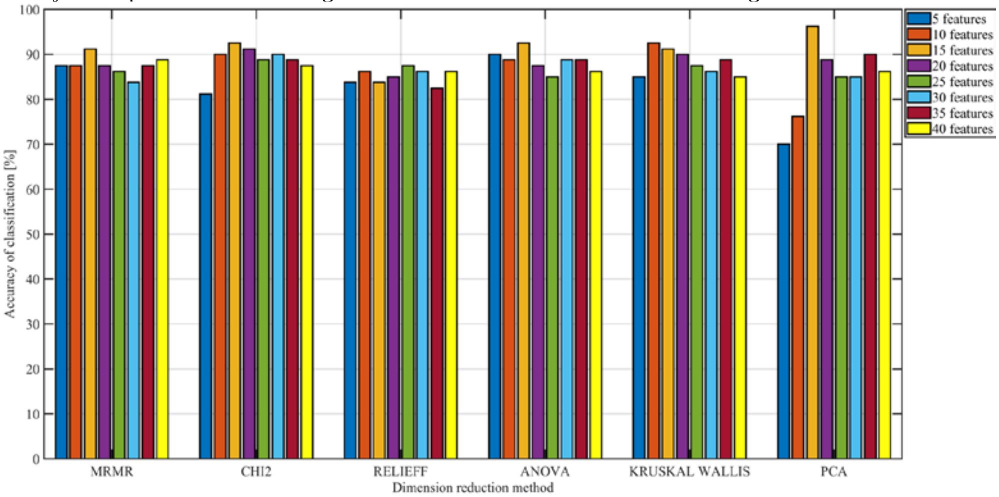

Fig. 3. LDA performances for different feature spaces

Figure 3 shows that the LDA classifier is more successful in the 15-dimensional space. In 4 out of 6-dimension reduction methods, the best performance was obtained in 15-dimensional space. The fluctuation in classification accuracies depending on the dimension of the feature space is highest for the PCA method.

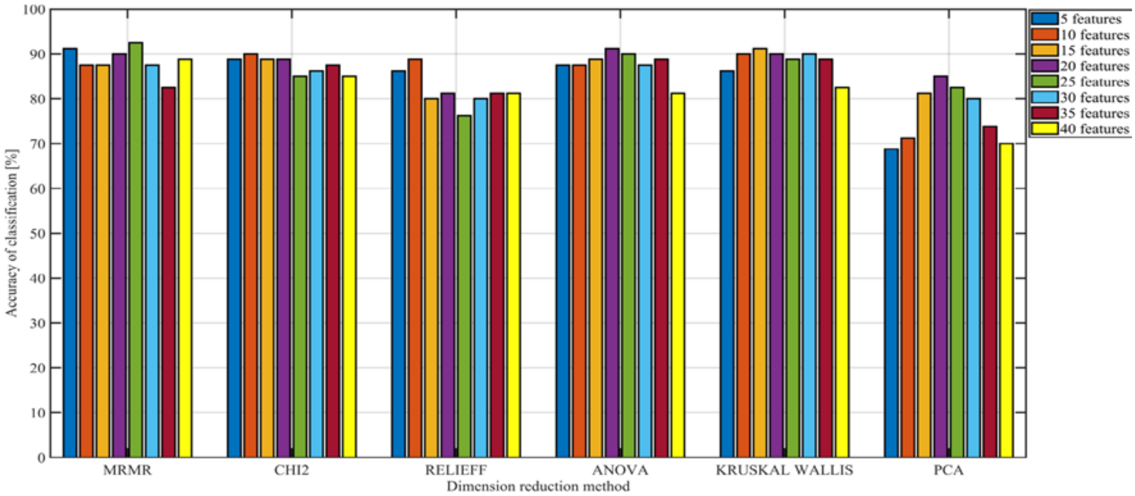

Fig. 4. SVM performances for different feature spaces

SVM classifier is more successful in 10 and 20-dimensional spaces. The classification accuracies fluctuate less in the CHI2 method. The fluctuate in classification accuracies according to the dimension of the feature space was the highest in the PCA method, similar to the other classifiers.

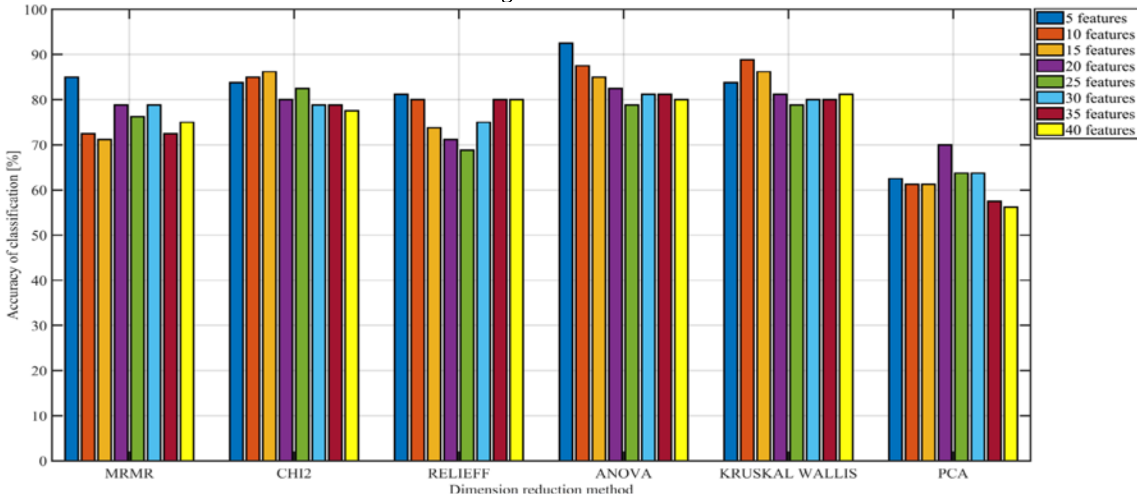

Fig.5. k-NN performances for different feature spaces

The k-NN classifier performed better in low dimensional spaces. In 3 of the 6 methods, the highest accuracy was obtained in the 5-dimensional feature space. With the LDA classifier, the best performance was achieved with 96.2% accuracy in the 15-dimensional space obtained with the PCA method. The lowest classification accuracy was 56.2% with PCA and k-NN methods in 40-dimensional space. The confusion matrices for these cases are given in Table 3 and Table 4.

Table 3. The confusion matrix for the best performance

		predicted			
		Relax	Physical	Cognitive	Emotional
true	Relax	20			
	Physical		20		
	Cognitive		2	17	1
	Emotional				20

Table 4. The confusion matrix for the worst performance

		predicted			
		Relax	Physical	Cognitive	Emotional
true	Relax	15	1	2	2
	Physical	3	10	4	3
	Cognitive	6	2	9	3
	Emotional	3	1	5	11

The effects of the feature space on the stress detection performance are clear from these two tables. Classification accuracy increases when the appropriate feature space is created. Even at its best, the classifier struggled to accurately detect cognitive stress. Three samples of cognitive stress were confused with physical and emotional stress. Relaxation state, physical stress state and emotional stress state were detected without error.

4. Conclusions

Diagnosing neurological conditions is vital in treating anxiety and depression. Technologies that perform the function of monitoring neurological status without affecting daily life are being designed. In this way, a database can be created about the mental health of individuals. Feature sets created from the obtained data are analyzed in detail with machine learning methods. It is desired that the feature space carries information describing the neurological state. For this purpose, unnecessary features are eliminated. The findings obtained in the study show that the detection accuracy varies depending on the dimension of the feature space. The LDA classifier was found to be the most successful method. ANOVA, KW, and CHI2 performances were the most notable among the dimension reduction methods. Although PCA is the dimension reduction method with the highest accuracy, it was the method with the highest fluctuation in classification accuracies depending on the feature set dimension.

Declaration of conflicting interests

The authors declare no competing interests.

Funding

The author received no financial support for the research and/or authorship of this article.

References

[1] Kajantie, E., & Phillips, D. I. "The effects of sex and hormonal status on the physiological response to acute psychosocial stress." *Review*,2006.

[2] Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. "Review on psychological stress detection using biosignals." *IEEE Transactions on Affective Computing*, 13(1), 440-460,2019.

[3] Birjandtalab, J., Cogan, D., Pouyan, M. B., & Nourani, M. "A non-EEG biosignals dataset for assessment and visualization of neurological status."*In 2016 IEEE International Workshop on Signal Processing Systems (SiPS)* (pp. 110-114). IEEE,2016.

[4] Priya, A., Garg, S., & Tigga, N. P. " Predicting anxiety, depression and stress in modern life using machine learning algorithms." *Procedia Computer Science*, 167, 1258-1267,2020.

[5] Pham, T. "Visualization and Classification of Neurological Status with Tensor Decomposition and Machine Learning" *Digitala Vetenskapliga Arkivet.* 2019.

[6] Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. "Automatic driver stress level classification using multimodal deep learning." *Expert Systems with Applications*, 138, 112793,2019.

[7] Arpaia, P., Moccaldi, N., Prevete, R., Sannino, I., & Tedesco, A. "A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis." IEEE Transactions on Instrumentation and Measurement, 69(10), 8335-8343,2020.

[8] Bobade, P., & Vani, M. "Stress detection with machine learning and deep learning using multimodal physiological data." *In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 51-57). IEEE, 2020.

[9] Jaloli, M., Choudhary, D., & Cescon, M. "Neurological Status Classification Using Convolutional Neural Network." *IFAC-PapersOnLine*, 53(5), 409-414,2020.

[10] Masood, S., Khan, R., Abd El-Latif, A. A., & Ahmad, M. "An FCN-LSTM model for neurological status detection from non-invasive multivariate sensor data." *Neural Computing and Applications*, 1-17,2022.

[11] Iqbal, T., Elahi, A., Wijns, W., & Shahzad, A. "Exploring unsupervised machine learning classification methods for physiological stress detection." *Frontiers in Medical Technology*, 4, 782756,2022.

[12] *A non-EEG dataset for assessment of neurological status*, PhysioNet. [On-line]. Available on: https://archive.physionet.org/physiobank/database/noneeg/, Accessed on: August 2023.

[13] Gersch, W. "Estimation of the autoregressive parameters of a mixed autoregressive moving-average time series." *IEEE Transactions on Automatic Control*, 15(5), 583-588,1970.

[14] Bos, R., De Waele, S., & Broersen, P. M. "Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data." *IEEE Transactions on Instrumentation and Measurement*, 51(6), 1289-1294,2002.

[15] Özyurt, F. "A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine." *Soft Computing*, 24(11), 8163-8172,2020.

[16] Liu, H., & Setiono, R. "Chi2: Feature selection and discretization of numeric attributes." *In Proceedings of 7th IEEE international conference on tools with artificial intelligence*, 388-391,1995.

[17] Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. "Relief-based feature selection: Introduction and review." *Journal of biomedical informatics*, 85, 189-203,2018.

[18] Ali Khan, S., Hussain, A., Basit, A., & Akram, S. "Kruskal-Wallis-based computationally efficient feature selection for face recognition." *The Scientific World Journal*, 2014.

[19] Bro, R., & Smilde, A. K. "Principal component analysis." Analytical methods, 6(9), 2812-2831,2014.

[20] Balakrishnama, S., & Ganapathiraju, A. "Linear discriminant analysis-a brief tutorial." *Institute for Signal and information Processing*, 18(1998), 1-8,1998.

[21] Burges, C. J. "A tutorial on support vector machines for pattern recognition." *Data mining and knowledge discovery*, 2(2), 121-167,1998.

[22] Kataria, A., & Singh, M. D. "A review of data classification using k-nearest neighbour algorithm." *International Journal of Emerging Technology and Advanced Engineering*, 3(6), 354-360,2013.