Stress Detection in EEG Signals Using SVM Algorithm and Spectral Representations

Alfredo Lapoint Angulo[1] Sergio Renato Rengifo Lozano[2]
Cristian Jesús Echejaya Moreno[1] Yarol Sebastian Nolasco Guia[1]

[1]National University of San Marcos, Faculty of Electronic Engineering
Av. Carlos Germán Amezaga 375, Cercado de Lima, Peru
[2]National University of Trujillo

Correspondence: alfredo.lapoint@unmsm.edu.pe (A.L.A.), P880301324@unitru.edu.pe (S.R.R.L.), cristian.echejaya@unmsm.edu.pe (C.J.E.M.), yarol.nolasco@unmsm.edu.pe (Y.S.N.G.)

Abstract

Stress is an essential physiological response that, under prolonged conditions, can become a significant risk to mental and physical health. Given the limitations associated with traditional assessment methods, there is a need to implement automatic, objective, and reproducible systems that allow accurate stress detection. This study proposes the design and implementation of a classifier based on Support Vector Machines (SVM) to identify binary states of stress and relaxation from EEG signals. The methodology includes a rigorous preprocessing pipeline, multiresolution decomposition using Discrete Wavelet Transform, spectral analysis with PSD, and extraction of statistical features. To preserve the nonlinear and non-stationary nature of EEG signals, techniques that minimize the loss of relevant physiological information are employed. The results are expected to validate the feasibility of SVM models as a diagnostic support tool in clinical and technological contexts, maintaining high accuracy levels without compromising neural data integrity.

Keywords: stress, EEG signals, SVM, PSD, automatic classification

Introduction

Stress has been defined by the World Health Organization (WHO) as a state of worry or mental tension caused by difficult situations [1]. Although it constitutes an adaptive response of the organism to challenges or threats, its prolonged or dysregulated manifestation has become a relevant public health problem, given its high prevalence and negative impact on individuals' physical and psychological well-being.

This growing relevance as a clinical phenomenon has highlighted the limitations of traditional assessment methods, which mainly depend on subjective self-reports and clinical observations susceptible to interpretive biases. In this context, the need to establish more objective and standardized assessment methods has motivated various investigations, such as the preliminary studies developed by [2], which highlight the importance of reproducible tools for accurate stress detection.

From this approach, recent research such as that by [3] has turned to bioengineering techniques to identify stress states from the analysis of biosignals, particularly electroencephalographic (EEG) signals. These signals, when processed through advanced mathematical methods and artificial intelligence models, allow the development of classification systems capable of discriminating between stressed and non-stressed states. Among the most used machine learning algorithms in this field are K-Nearest Neighbors (KNN), logistic regression, Support Vector Machines (SVM), and Random Forest [4].

Study Approach

Although various approaches have shown promising results when applying signal processing techniques combined with hybrid machine learning frameworks to classify subjects' stress state [5], their clinical application still presents limitations. In particular, to be considered viable in real clinical environments, these models must achieve accuracy levels above 95% with error margins below 5% [6].

In the study by [7], a Support Vector Machine (SVM) is trained using, as part of its methodology, a "data augmentation" criterion aimed at mitigating the influence of outliers present in the EEG dataset. This technique consists of expanding the original samples by scaling artificial signals using preselected factors close to unity. However, according to [8], this strategy can induce biases in the dataset, thus compromising the veracity of the SVM classifier's accuracy metrics.

This challenge highlights the need to continue developing methodologies that not only optimize classification performance but also respect the inherent complexity of brain biosignals, characterized by their highly stochastic and non-stationary behavior.

Research Problem

How to implement an SVM classifier that accurately detects stress from EEG signals without altering their nonlinear and non-stationary complexity?

Justification

Clinical: From a clinical perspective, stress diagnosis requires more objective methods than traditional questionnaires, which can be subjective and inconsistent. The use of EEG signals allows access to direct physiological information about brain state, offering a more reliable and quantifiable pathway for stress detection.

Technical: From a technical perspective, EEG signals possess considerable complexity due to their nonlinear, stochastic, and non-stationary nature. This structure demands the use of robust classifiers capable of preserving these properties during processing. In this context, Support Vector Machines (SVM) are particularly suitable due to their ability to handle high-dimensional data and perform accurate classifications even with moderate data volumes.

General Objective

Implement and evaluate an SVM classifier to detect stress from EEG signals, respecting the inherent complexity of such signals and maximizing model accuracy.

Specific Objectives

1. Preprocess EEG signals to preserve their nonlinear and non-stationary properties, removing noise without losing relevant information.

2. Extract representative features from EEG signals that effectively distinguish between stress and non-stress states.

3. Train and validate an SVM model using the extracted features, optimizing its parameters to achieve high classification accuracy.

Methodology

Research Design

The project will adopt a quantitative and experimental approach, oriented towards the development and evaluation of an SVM classifier that will be trained and validated using pre-existing data, under a controlled scheme that will allow objective evaluation of its accuracy and preservation of the inherent characteristics of EEG signals. The methodology will be structured in consecutive stages of preprocessing, feature extraction, and classification, with quantitative measurement of performance.

Variable Operationalization

Table 1: Variable operationalization

Variable	**Indicator**	**MeasurementTechnique**
Classifier accuracy	Percentage of correct classifications of stress vs. relaxation	Accuracy, precision, recall, F1-score
Preservation of EEG signal randomness and non-stationarity	Degree of conservation of original dynamic properties after preprocessing	Measurement through spectral entropy before, during, and after processing

Data Source

EEG recordings from the open repository "EEG during mental arithmetic tasks" of PhysioNet will be used [9]. The signals were recorded with the 23-channel Neurocom system according to the international 10/20 system, and are available in EDF format. Each subject has two recordings: one in a baseline state (relaxed) and another during a mental arithmetic task (stressed). The dataset includes 36 participants, divided into two groups according to task performance: Group G (high performance) and Group B (low performance). Binary classification will be established to discriminate between stress and relaxation states. An 80/20 partition criterion will be used to separate the data into training and test sets.

EEG Signal Preprocessing

After downloading the data corresponding to stress and relaxation states, a preliminary inspection was carried out to evaluate the sample distribution in each set. Considerable imbalance was identified. The relaxation state contains 60,002 samples per participant, while the stress state contains 30,001 samples per participant. Both signals were recorded with a sampling frequency of 500 Hz.

To balance both sets and avoid bias in the model training phase, a scaling technique will be applied to the signals of the minority set. This will consist of generating new samples by multiplying the original signals by a random factor between 0.97 and 1.03, thus preserving the statistical characteristics and randomness of the original signal. Finally, both sets will be equalized to a total of 90,001 samples per participant for each set.

Multiresolution Decomposition

To address the non-stationary and highly complex nature of EEG signals, the Discrete Wavelet Transform (DWT) will be used with the Daubechies order 4 (db4) function as the mother wavelet. This choice is supported by the properties that this wavelet offers for multiresolution decomposition, such as its smoothness and compact support, which will allow capturing transients and local patterns in the signal without introducing significant distortions [10].

The application of DWT will decompose each channel into detail coefficients (CD) and approximation coefficients (CA) up to a total of 8 levels. This decomposition will facilitate the effective separation of high and low frequency components, preserving the time-frequency structure of the signal, fundamental for preserving its inherent complexity.

Power Spectral Density

Once the approximation and detail coefficients are obtained through the Discrete Wavelet Transform (DWT), the Fast Fourier Transform (FFT) will be applied to analyze their spectral distribution. This stage will allow identifying the dominant components in the frequency domain, thus complementing the multiresolution analysis with a more precise frequency perspective.

The application of FFT on DWT coefficients will allow analyzing the energy distribution in the frequency domain through power spectral density (PSD). This metric is commonly used to identify relevant patterns in EEG signals and detect possible anomalies associated with different levels of brain activation [11].

Feature Extraction

Once the histogram of the power spectral density (PSD) is obtained, the extraction of statistical features will proceed, based on the approach proposed by [12]. The selected features will capture different properties of the signal's spectral distribution and are commonly used in EEG signal analysis for their physiological relevance and computational robustness. The extracted metrics will be as follows:

- **Median (x):** central value of the power distribution, robust against outliers.

$$x_{(n+1)/2} \quad \text{if } n \text{ is odd} \tag{1}$$

- **Root Mean Square (RMS):** will estimate the average signal energy, being sensitive to high magnitude values.

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{2}$$

- **Kurtosis ():** will measure the concentration of energy around the mean, useful for detecting the presence of peaks or anomalous events.

$$\kappa = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 \tag{3}$$

  where $\bar{x}$ is the mean and σ the standard deviation.

- **Norm ($\|x\|_2$):** will represent the total magnitude of the signal in Euclidean space.

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{4}$$

- **Entropy (H):** will quantify the complexity or randomness of the power distribution, useful for detecting chaotic or structured signals.

$$H(x) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{5}$$

  where p_i represents the normalized probability of the value x_i within the histogram.

These metrics will be calculated on each PSD histogram generated from the coefficients obtained in the DWT stage, thus enriching the feature vector that will later feed the classification model.

State Classification

For binary classification of stress and relaxation states, an SVM classifier will be used, trained using different kernels: linear, radial basis function (RBF), polynomial, and sigmoidal. To maximize model performance, the PSO algorithm will be applied for hyperparameter optimization.

PSO is a population-based optimization technique inspired by the collective behavior of swarms, which allows efficient search in high-dimensional solution spaces [13]. In this context, the parameters to optimize include the penalty coefficient C, the γ parameter in nonlinear kernels, and the polynomial degree in the polynomial kernel.

The SVM model seeks to find the optimal hyperplane that maximizes the margin between classes, which can be formalized through the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \quad \text{subject to} \quad y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{6}$$

where w is the weight vector, b is the bias, ξ_i are the slack variables that allow some tolerance to classification errors, C controls the balance between model complexity and error, and $\phi(\cdot)$ represents the mapping function to the feature space defined by the kernel.

Performance Evaluation

The classifier performance will be evaluated using the following standard metrics:

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN}$
- **Precision:** $\frac{TP}{TP+FP}$
- **Recall:** $\frac{TP}{TP+FN}$
- **F1-score:** $2 \times \frac{Precision \times Recall}{Precision+Recall}$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Preliminary Results

To avoid overloading this article with all individual graphical results per subject, subject No. 10 and channel Fp2 have been selected as representative examples, which has shown one of the highest activations during the execution of complex tasks under time constraints [14]. The complete results, including graphs, tables, and source code, are available in the GitHub repository indicated in the Annexes section.

Entropy Analysis

An analysis of entropy was carried out in three successive instances, to capture the informational complexity of the signal at different processing stages:

1. **Coefficient entropy:** Calculated directly on the coefficients obtained through the discrete wavelet transform (DWT) applied to the EEG signal of channel Fp2.
2. **FFT entropy:** Calculated on the spectra generated by applying the fast Fourier transform (FFT) to each set of DWT coefficients.
3. **PSD entropy:** Calculated from the power spectral distributions obtained through the Welch method applied to the FFT spectra (DWT + FFT + PSD).

It was observed that, in the coefficients corresponding to low frequencies (CA8 to CD4), subjects in the relaxation condition consistently present higher entropy levels than those in the stress condition. This difference is attributed to the fact that power distributions in stressed subjects tend to be more uniform or flat, that is, with fewer abrupt transitions, which reduces the measured entropy. In contrast, relaxed subjects show greater variability in power distribution, reflected in higher entropy levels.

Classifier Performance

The classification model was trained using four different kernel types, optimizing their respective hyperparameters through the Particle Swarm Optimization (PSO) algorithm. The results obtained after this process are summarized in Table 1.

Table 2: SVM classifier performance according to kernel type

Kernel	**C**	**Gamma**	**Degree**	**Accuracy**	**F1**	**Prec.**	**Recall**
Linear	16.68	scale	–	0.999	0.999	0.999	0.999
RBF	85.61	0.0001	–	0.999	0.999	0.999	0.999
Sigmoid	10.46	0.0001	–	0.999	0.999	0.999	0.999
Polynomial	3.98	0.3067	3	0.790	0.730	0.999	0.57

Discussion

Experimental results indicated accuracy above 99% for the different kernel types used (linear, RBF, polynomial, and sigmoidal), highlighting the robustness and capacity of this model to handle binary classification accurately. A key aspect of the analysis was the use of entropy to differentiate between stress and relaxation states. It was observed that subjects in a stressed state presented significantly lower entropy levels, indicating a more uniform power distribution and less variability in EEG signals. In contrast, relaxed subjects showed higher entropy levels, reflecting greater variability in brain signals.

The approach adopted in preprocessing EEG signals, using the Discrete Wavelet Transform (DWT), allowed preserving the nonlinear and non-stationary properties of the signals, which is essential to avoid losing relevant physiological information during processing. This step proved crucial to ensure that the model could work with signals that maintain their inherent complexity, fundamental for classifier accuracy.

The optimization of classifier hyperparameters using Particle Swarm Optimization (PSO) also had a positive impact on model performance, achieving almost perfect performance metrics, such as precision, recall, and F1-score with values above 0.99 compared to the metrics obtained by [7]. This optimization allowed finding the best balance between model complexity and performance, ensuring consistent results.

Despite these achievements, the study also highlighted some limitations related to the presence of noise and artifacts in EEG signals. These factors can affect model accuracy, so it is essential to perform rigorous preprocessing to mitigate these effects. Additionally, the use of data augmentation techniques can introduce biases if not properly managed, which should be considered in future research.

Conclusions

In this study, it was demonstrated that the classifier based on Support Vector Machines (SVM) is a highly effective tool for stress detection in EEG signals, achieving outstanding results in the classification of stress and relaxation states. Experimental results indicated accuracy above 99% for the different kernel types used, highlighting the robustness and capacity of this model to handle binary classification accurately.

A key aspect of the analysis was the use of entropy to differentiate between stress and relaxation states. It was observed that subjects in a stressed state presented significantly lower entropy levels, indicating a more uniform power distribution and less variability in EEG signals. In contrast, relaxed subjects showed higher entropy levels, reflecting greater variability in brain signals. This difference in entropy levels has been established as a fundamental characteristic for state classification.

The approach adopted in preprocessing EEG signals, using the Discrete Wavelet Transform (DWT), allowed preserving the nonlinear and non-stationary properties of the signals, which is essential to avoid losing relevant physiological information during processing. This step proved crucial to ensure that the model could work with signals that maintain their inherent complexity, fundamental for classifier accuracy.

The optimization of classifier hyperparameters using Particle Swarm Optimization (PSO) also had a positive impact on model performance, achieving almost perfect performance metrics. This optimization allowed finding the best balance between model complexity and performance, ensuring consistent results.

In terms of applications, the SVM model trained in this study can offer a valuable tool for objective stress detection, with potential applications in both clinical and technological environments. The use of EEG signals provides a direct and quantifiable way to access the physiological state of the brain, representing a significant improvement over traditional methods based on subjective self-reports. This suggests that this approach could contribute to better assessment and diagnosis of stress in various areas, such as mental health, medicine, and occupational well-being.

Acknowledgments

We thank the National University of San Marcos for the institutional support provided for this research work.

Declaration of Interests

The authors declare no conflicts of interest.

Annexes

Code Repository

The complete source code, including graphs, tables, and scripts used in this study, is available in the following GitHub repository:

`https://github.com/Alfredo-Lapoint/BrainwaveAI-StressDetector`

References

[1] World Health Organization. ”Stress.” (2025), available: `https://www.who.int/news-room/questions-and-answers/item/stress` (accessed 05-06-2025).

[2] K. Kiuchi, X. Kang, R. Nishimura, M. Sasayama and K. Matsumoto, ”Predicting physical and mental health status through interview-based evaluation of work stress: initial attempts to standardize the interviewing method,” *Industrial Health*, vol. 62, no. 4, pp. 237-251, 2024.

[3] Y. Badr, U. Tariq, F. Al-Shargie, F. Balbioni, F. Al Mughairbi and H. Al-Nashash, ”A review on evaluating mental stress by deep learning using EEG signals,” *Neural Computing and Applications*, vol. 36, no. 21, pp. 12 629-12 654, 2024.

[4] S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," *IEEE Access*, vol. 9, pp. 84 045-84 066, 2021.

[5] S. S. Hosseini, M. Yamaghani and S. Poorzaker Arabani, "A review of methods for detecting multidimensional emotions in sound, image and text," *Journal of Applied Research on Industrial Engineering*, vol. 11, no. 4, pp. 669-705, 2024.

[6] R. Fernandez-Matias, "Sample Size Calculation in Health Sciences: Recommendations and Practical Guide," *Journal of MOVE and Therapeutic Science*, vol. 5, no. 1, pp. 481-503, 2023.

[7] H. Goker, *Comparison of different spectral analysis methods with an experimental EEG dataset*, 2022.

[8] A. Sakai, Y. Minoda and K. Morikawa, "Data augmentation methods for machine-learning-based classification of bio-signals," in *2017 10th Biomedical Engineering International Conference (BMEICON)*, 2017.

[9] I. Zyma, S. Tukaev and I. Seleznov, *EEG during mental arithmetic tasks*, `https://physionet.org`, Accessed: 2025-06-05, 2018.

[10] A. Rafiki et al., "Implementation of Vision Transformer for early detection of autism based on EEG signal heatmap visualization," *International Journal of Electrical, Electronics and Mechatronics Engineering (IJEEEMI)*, vol. 7, no. 1, pp. 102-112, 2025.

[11] R. Wang, J. Wang, H. Yu, X. Wei, C. Yang and B. Deng, "Power spectral density and coherence analysis of Alzheimer's EEG," *Cognitive Neurodynamics*, vol. 9, no. 3, pp. 291-304, 2015.

[12] K. AlSharabi, Y. Bin Salamah, A. M. Abdurraqeeb, M. Aljalal and F. A. Alturki, "EEG signal processing for Alzheimer's disorders using discrete wavelet transform and machine learning approaches," *IEEE Access*, vol. 10, pp. 89 781-89 797, 2022.

[13] M. Daviran, A. Maghsoudi and R. Ghezelbash, "Optimized AI-MPM: Application of PSO for tuning the hyperparameters of SVM and RF algorithms," *Computers & Geosciences*, vol. 195, no. 105785, p. 105 785, 2025.

[14] B. Gonzalez-Gomez, M. I. Nunez-Pena and Colome, "Math anxiety and the shifting function: An event-related potential study of arithmetic task switching," *European Journal of Neuroscience*, vol. 57, no. 11, pp. 1848-1869, 2023.