DSCP4WK - Data Science Capstone Project

Problem 18: EEG Based Mental Stress Detection

Authors Leen AL MAJZOUB
Anubhav LAMSAL
Botond HERNYES

Lecturers Dr. Dipl.-Ing. Stefan BISCHOF
Himanshu BUCKCHASH, PhD

Project Report
June 25, 2025

Abstract

This report demonstrates an effective EEG-based stress detection system using machine learning. Evaluating multiple models on 22 subjects' data, we found Neural Networks (96.1% accuracy) and Random Forest (92.8%) outperformed other methods. Hybrid features (spectral + quality metrics) and PCA-based dimensionality reduction proved critical for performance. The results validate EEG's utility for real-world stress monitoring applications.

Keywords: EEG, stress detection, machine learning, neural networks, random forest, hybrid features

Individual Contribution

Table 1 **Summary of Individual Contributions**

Group Member	Contribution Details
Anubhav (34%)	Preprocessing time-series data (came up with the pipeline to convert measurements into bands of frequency using Welch's). Designed the methodologies and testing suites (normality check, global differences, and pairwise comparison of data for correlation checks). Prepared the initial parameters and pipeline setup for ML modeling (to be refined by Botond) and applied dimensionality reduction for yielding the better results in the later stages.
Botond (33%)	Conducted initial stages of EDA and data understanding (checking for missing values, distribution plots, signal plots and so on). Utilized the pipelines created and ran experiments for parameter selections and yielded the best classification model. Conducted the literature review and authored the Results and Discussion sections.
Leen (33%)	Created the presentation slides, report template, and wrote the Abstract and Conclusion. Performed some preliminary literature review and data understanding.

Contents

1 Introduction

1.1 Problem Statement

We aim to investigate whether EEG-based features—specifically frequency-domain band-power, signal quality metrics, and sensor-level measures—can effectively differentiate between various emotional and cognitive scenarios, including high-stress conditions, relaxed states, and cognitively demanding problem-solving tasks.

1.2 Objectives

The goal is to develop a reproducible computational pipeline that extracts frequency-domain bandpower (δ, α, β, γ) and signal-quality metrics (e.g., signal-to-noise ratio, artifact rejection). The pipeline will be validated by ensuring bandpower estimates closely match Welch's method ($\leq$5% mean absolute error) and that quality metrics identify at least 95% of bad segments. Additionally, we will evaluate whether these EEG features can distinguish between high-stress and relaxed states by training a binary classifier. Success requires $\geq$80% average accuracy and ROC AUC $\geq$0.85, with performance significantly above chance ($p < 0.05$).

1.3 Background

EEG (electroencephalogram) measures the brain's electrical activity using a helmet with metal discs while subjects respond to various stimuli, such as solving math problems or watching movies, to observe brain reactions [1]. Brain activity is categorized into frequency bands: delta, theta, alpha, beta, and gamma, each linked to specific mental states and brain regions. Delta waves (1–4 Hz) relate to deep sleep and unconsciousness in the frontal cortex; theta waves (4–8 Hz) to relaxation, drowsiness, and working memory in frontal and temporal areas; alpha waves (8–13 Hz) to calm wakefulness in occipital and parietal regions; beta waves (13–30 Hz) to active thinking and cognitive load in the frontal cortex; and gamma waves (30–45+ Hz) to high-level cognition and attention across the brain, often in frontal and sensory areas [2].

1.4 Dataset

The dataset, collected using the EMOTIV EEG 5-Channel Sensor kit, consists of EEG recordings from 22 subjects during five types of stimulation: complex mathematical problem solving, Trier mental challenge, Stroop color word, horror video, and relaxing music.

Table 1.1 Summary of EEG Data Attributes

Field(s)	**Unit**	**Description**
Timestamps & Counters		
`TIME_STAMP_s, _ms, OR_TIME_STAMP_s, _ms`	s / ms	Primary and original device timestamps in seconds and milliseconds.
`COUNTER`	index	Sample index for sequence tracking and detecting drops.
`INTERPOLATED`	binary	Indicates missing or interpolated data (1 = interpolated).
EEG Channels (µV)		
`AF3, AF4, T7, T8, Pz`	µV	EEG voltage signals from frontal, temporal, and parietal electrodes.
Signal Quality & Battery		
`RAW_CQ`	a.u.	Real-time contact quality of EEG electrodes.
`BATTERY, BATTERY_PERCENT`	a.u. / %	Battery level (raw value and percentage).
Markers (Events)		
`MarkerIndex, MarkerType, MarkerValueInt, MARKER_HARDWARE`	a.u.	Event markers including ID, type, associated code, and source (hardware/software).
Contact Quality (CQ)		
`CQ_AF3, CQ_AF4, CQ_T7, CQ_T8, CQ_Pz, CQ_OVERALL`	a.u.	Electrode-specific and overall contact quality (0 = bad, 1 = good).
EEG Signal Quality (EQ)		
`EQ_SampleRateQua, EQ_AF3, EQ_AF4, EQ_T7, EQ_T8, EQ_Pz, EQ_OVERALL`	a.u.	Signal quality per electrode and overall; includes sample rate stability.

2 Literature Review

EEG has been researched consistently over the past two decades, from all over the world to find meaning behind the different values of brain activities. We have looked at multiple research papers that investigated the relation of EEG features and different stress levels. One interesting read was a research done in [3], where they looked at and compared different machine learning models from prior research. The study evaluated these models based on accuracy, publication year, number of participants and EEG channels used, experimental conditions (e.g., relaxed vs. stressed states), and other relevant factors. The models have shown a promising average of 89.97% for the older models and 85.36% for the newer models. These were selected based on the conditions, all of the models included in the average had a “relaxed vs. some levels of stress” condition, which is the same category in which we set out to make our models.

A different study [4] compared different machine learning methodologies to see which one is the most effective when predicting EEG data. After looking at the results of several models with same and different methods, the researchers found that support vector machine (SVM), neural network, and random forest models produce consistently and significantly better results than the rest of the models with different methods.

Another study [5] we looked at solved the issue of detecting stress from EEG data with a hybrid approach. They took the normal EEG data (brain activity on different frequencies) and integrated time as another variable into the model. They compared a frequency domain only model, a time domain only model and a hybrid model. The latter showed the highest accuracy with 95% for the random forest model and 98.33% for the SVM model.

With these information in mind, we set out to approach our problem in a similar manner, using multiple methods of machine learning with regular and hybrid data to create a model that can hopefully predict with a higher accuracy than the researched models from the first paper.

3 Methodology

3.1 Segmentation via Event Markers

In our paradigm, digital markers [6] embed the onset and offset of each experimental block into designated channels. Let $\{t_k\}$ be the sequence of nonzero marker timestamps. We define the sth segment as the interval $[t_k,\, t_{k+1}]$. By cropping the raw EEG signal $x_c(t)$ on channel c to these intervals, $x_c^{(s)}(t) \;=\; x_c(t)\Big|\, t_k \le t \le t_{k+1},$, we make sure that each segment reflects a homogeneous brain-state epoch, minimizing transitional artifacts.

3.2 Welch's Method for Power Spectral Density

We extracted spectral band-power features from each EEG segment and saved them in as an aggregated CSV. Denote by $x_c(t)$ the raw voltage time-series (in volts) recorded on channel $c \in \{\mathrm{AF3}, \mathrm{T7}, \mathrm{Pz}, \mathrm{T8}, \mathrm{AF4}\}$ [7], sampled at $f_s = 128\,\mathrm{Hz}$. Let $\{\, t_k\}$ be the nonzero marker times that define segment boundaries. For the s-th segment $[t_k, t_{k+1}]$, we denote the discrete, unit-converted signal by

$$x_c^{(s)}[n] \;=\; 10^6\, x_c\Big(t_k + n/f_s\Big), \quad n = 0, \dots, N_s - 1,$$

where $N_s = \lfloor (t_{k+1} - t_k)\, f_s \rfloor$.

Welch PSD estimation [8] Each segment $x_c^{(s)}[n]$ was divided into K overlapping subsegments of length $N_{\mathrm{seg}} = 256$ samples (default 50% overlap) and windowed by a Hann window $w[n]$. The Welch power spectral density estimate at frequency $f_m = m\, f_s/N_{\mathrm{seg}}$ is

$$\widehat{S}_c^{(s)}(f_m) \;=\; \frac{1}{K\,U} \sum_{k=0}^{K-1} \Big|\, \mathcal{F}\{\, w[n]\, x_c^{(s)}[n + k\, N_{\mathrm{seg}}/2]\} \Big|^2$$

where $\mathcal{F}\{\cdot\}$ is the discrete Fourier transform and $U = \sum_{n=0}^{N_{\mathrm{seg}}-1} w^2[n]$ normalizes window energy.

Band-power computation Define the five classical EEG bands

$$\mathcal{F}_\Delta = \{\, f : 1 \le f < 4\}, \quad \mathcal{F}_\Theta = \{\, 4 \le f < 8\},$$
$$\mathcal{F}_\alpha = \{\, 8 \le f < 13\}, \quad \mathcal{F}_\beta = \{\, 13 \le f < 30\},$$
$$\mathcal{F}_\Gamma = \{\, 30 \le f < 45\}.$$

The mean power in band b for channel c in segment s is

$$P_{c,b}^{(s)} = \frac{1}{|\mathcal{F}_b|} \sum_{f_m \in \mathcal{F}_b} \widehat{S}_c^{(s)}(f_m), \quad b \in \{\Delta, \Theta, \alpha, \beta, \Gamma\}.$$

Auxiliary features Let $\{q_j(t)\}$ be the continuous channel-quality or electrode-quality time-series (columns `CQ_AF3`, ..., `EQ_OVERALL`) and $\{B(t)\}$ the battery level signals (`BATTERY`, `BATTERY_PERCENT`). For each segment s, we computed

$$\mu_{q_j}^{(s)} = \frac{1}{N_s} \sum_{n=0}^{N_s-1} q_j\Big(t_k + n/f_s\Big), \quad \sigma_{q_j}^{(s)} = \sqrt{\frac{1}{N_s} \sum_{n=0}^{N_s-1} \Big(q_j(n) - \mu_{q_j}^{(s)}\Big)^2},$$

and analogously $\mu_B^{(s)}, \sigma_B^{(s)}$.

3.3 Auxiliary Quality and Battery Metrics

To control for recording artifacts and device status, we also extracted per-segment means and standard deviations of:

- Channel-quality signals $\{q_j(t)\}$ (columns `CQ_`, `EQ_`),
- Battery level signals $B(t)$ (`BATTERY`, `BATTERY_PERCENT`).

For each metric m, we compute

$$\mu_m^{(s)} = \frac{1}{N_s} \sum_{n=0}^{N_s-1} m(t_k + n/f_s), \quad \sigma_m^{(s)} = \sqrt{\frac{1}{N_s} \sum_{n=0}^{N_s-1} \Big(m(t_k + n/f_s) - \mu_m^{(s)}\Big)^2}.$$

3.4 Statistical Tests and Conclusions

Table 3.1 Summary of Statistical Analyses on Band-Power Features

Stage	Test	Key Findings
Normality Check	Shapiro–Wilk	All 25 band-scenario distributions were non-normal ($p < 10^{-18}$); nonparametric methods used.
Global Differences	Kruskal–Wallis H	All frequency bands showed significant differences across scenarios ($p < 10^{-5}$).
Pairwise Comparison	Dunn's Test (Bonferroni)	Horror scenario differed from all others ($p_{\text{adj}} < 0.001$), Relaxing music had higher Theta/Alpha than cognitive tasks. Additional contrasts found among cognitive tasks.

EEG band-power patterns differ across mental states: horror videos increased high-frequency (Beta, Gamma) activity, indicating arousal; relaxing music boosted low-frequency (Theta, Alpha) power, reflecting calmness. Cognitive tasks showed intermediate profiles. These results support spectral band-power as a reliable marker for classifying mental states.

3.5 Dimensionality Reduction

After extracting thirty-seven band-power, quality and battery features, we observed substantial multicollinearity (e.g. Delta vs. Theta, Beta vs. Gamma) and a high overall dimensionality that could impede classifier performance. To address these issues, we applied Principal Component Analysis (PCA) to the standardized feature matrix. PCA transforms the original correlated features into a smaller set of uncorrelated components that capture the majority of variance in the data. By projecting onto these components, we both reduce noise and minimize redundancy, which simplifies downstream modeling and helps prevent overfitting.

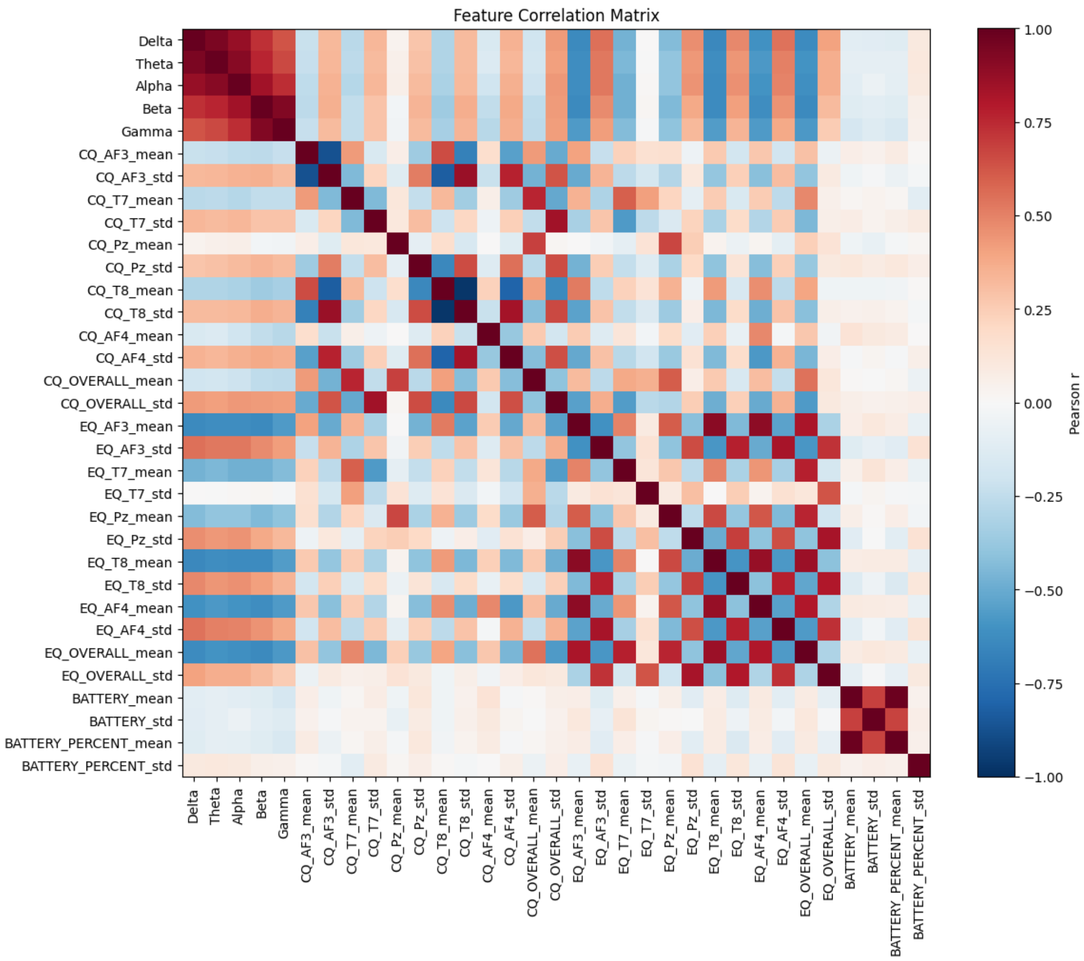

Figure 3.1 Correlation matrix of the different features [9], highlighting a clear correlation of several brain-related features

4 Results

Table 4.1 Summary of Model Performance on EEG Classification

Model	Accuracy (%)	Notes / Key Metrics
Support Vector Machine (RBF kernel) [10, 11]	66.3	Accuracy improved from 55.9% (10 components) to 66.3% (20 components). Lower performance, especially in distinguishing emotional vs. cognitive states.
Random Forest [12]	92.8	Trained on 20 PCA features + battery metrics; strong F_1 scores (0.96 Relaxing Music, 0.89 Horror Video). Marginal gains beyond 20 components.
XGBoost [13]	90.2	Achieved 81.0% accuracy on 10 components and 90.2% on 20. Outperformed SVM but below Random Forest. Consistent F_1 around 0.80 across classes.
Neural Network [14]	96.1	PyTorch feed-forward with ReLU; 100 epochs with Adam optimizer. 93.5% accuracy without battery features. Highest accuracy and F_1 scores (>0.94) among all models.

Model Comparison

Overall, ensemble tree methods (Random Forest, XGBoost) and deep neural networks markedly outperformed SVM. The Random Forest provided a strong baseline (92.8 %), while the neural network achieved the best accuracy (96.1 %), validating the efficacy of learned non-linear representations in distinguishing cognitive and emotional EEG states.

5 Discussion

The comparative evaluation of multiple classifiers on the PCA–reduced EEG feature set reveals several key insights. First, ensemble methods such as Random Forest and XGBoost consistently outperformed kernel-based SVM, indicating that decision-tree architectures better capture the complex, multi-modal relationships among spectral, quality, and battery features. Second, the deep neural network yielded the highest overall accuracy (96.1 %), suggesting that hierarchical, non-linear transformations further enhance discrimination of subtle differences in brain dynamics across emotional and cognitive states. Third, inclusion of battery metrics provided a modest but consistent performance boost, implying that device-level covariates can inform artifact-robust classification. Nevertheless, despite high test accuracies, our models were trained and evaluated on a single dataset collected under controlled conditions; external validation on independent cohorts and real-world scenarios (e.g. ambulatory monitoring) is needed to confirm generalizability. Future work should also explore temporal architectures (e.g. recurrent or convolutional neural networks) to leverage fine-grained time-series structure and assess the trade-off between model complexity and interpretability for practical neurotechnology applications.

6 Conclusion

This research shows a common approach and clear proof of how one can detect different stress levels in various situations with the use of EEG data. We successfully made several machine learning models with varying accuracy, the best ones being neural networks and random forest models. As our reference papers have pointed out, these methods perform significantly better than other machine learning methods. Our models have achieved an accuracy high enough for us to safely say that it is feasible to detect stress levels in different scenarios using EEG brain frequencies. These results, 92.8% for random forest and a notable 96.1% for neural network, performed higher than the average of the models collected from previous research papers. We also found that using hybrid data instead of only using brain frequency data significantly improved the accuracy of our models.

Bibliography

[1] Eráclito Niedermeyer and Fernando Lopes da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 5th. Lippincott Williams & Wilkins, 2005.

[2] György Buzsáki. *Rhythms of the Brain*. Oxford University Press, 2006.

[3] Rania Katmah et al. "A Review on Mental Stress Assessment Methods Using EEG Signals". In: *Sensors* 21.15 (2021), p. 5043. DOI: 10.3390/s21155043. URL: https://doi.org/10.3390/s21155043.

[4] Manoochehr Razavi et al. "Machine Learning, Deep Learning, and Data Preprocessing Techniques for Detecting, Predicting, and Monitoring Stress and Stress-Related Mental Disorders: Scoping Review". In: *JMIR Mental Health* 11.3 (2024), e53714. DOI: 10.2196/53714. URL: https://doi.org/10.2196/53714.

[5] Dr. Udesang K. Jaliya Ankita Gandhi. "Stress Detection through EEG Signals: Employing a Hybrid Approach integrating Time Domain, Frequency Domain Features and Machine Learning Techniques". In: *J. Electrical Systems* 20.3 (2024), pp. 3965–3973. DOI: 10.52783/jes.8087. URL: https://doi.org/10.52783/jes.8087.

[6] Steven J Luck. "An Introduction to the Event-Related Potential Technique". In: *MIT Press* (2014).

[7] Herbert H. Jasper. "The ten–twenty electrode system of the International Federation". In: *Electroencephalography and Clinical Neurophysiology* 10.2 (1958), pp. 371–375.

[8] Peter D Welch. "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), pp. 70–73.

[9] John D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95.

[10] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine Learning* 20.3 (1995), pp. 273–297.

[11] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[12] Leo Breiman. "Random forests". In: *Machine Learning* 45.1 (2001), pp. 5–32.

[13] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2016, pp. 785–794.

[14] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019, pp. 8024–8035.