SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Machine Learning

COURSE: IT3190E

MENTAL STRESS DETECTION USING EEG HEADSET

Author

Nguyen Tri Thanh	20225457
Nguyen Xuan Thanh	20225460
Nguyen Hoang Son	20225525
Le Nhat Quang	20225522
Phan Tran Viet Bach	20225435
Nguyen Quoc Thai	20225456

Mentor

THAN QUANG KHOAT

June 2024

MENTAL STRESS DETECTION USING EEG HEADSET

Nguyen Tri Thanh, Nguyen Xuan Thanh, Nguyen Hoang Son, Le Nhat Quang, Phan Tran Viet Bach, Nguyen Quoc Thai

Abstract: Mental stress is one of the major contributors to a variety of health issues which have been created by neurologists and psychiatrists to determine the intensity of mental stress in its early phases. The key candidate chosen is the electroencephalogram (EEG) signal which contains valuable information regarding mental states and conditions. The goal of this study is to measure emotional responses using carefully selected music videos (MV) pieces that induce emotional stress states. EEG signal is collected from 5 subjects to form a dataset with the length of 30 minutes in total by Mindlink EEG headset. The dataset will then be categorised using Multilayer Perceptron (MLP), Decision Tree (DT), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). Accuracy, precision, recall, and F1-Score are used to assess the data's performance. Finally, Artificial Neural Network (ANN) had the best performance where ANN have accuracy 99%, precision 99%, recall 99% and F1-score 99% among all machine learning classifiers.

1. INTRODUCTION

Stress, a pervasive and detrimental condition, is intrinsically linked to various emotional states. It is widely recognized that stress encompasses a spectrum of emotional fluctuations, with significant changes in emotions often correlating with alterations in brainwave patterns. Understanding and accurately measuring these changes are crucial for developing effective interventions and technologies to manage stress and enhance mental well-being.

Numerous studies have demonstrated the feasibility of classifying emotions through sophisticated systems such as Brain-Computer Interfaces (BCIs), which utilize Electroencephalography (EEG) to monitor brainwave activity. These systems have shown promise in detecting and interpreting the neural correlates of emotional states, providing valuable insights into the dynamics of stress and emotion. However, the complexity, high cost, and cumbersome nature of these multi-channel EEG systems present significant barriers to their widespread adoption and practical application.

In light of these challenges, this research aims to develop an advanced artificial intelligence (AI) model capable of classifying emotions using a single-channel EEG headset. This approach seeks to simplify the technology while maintaining accuracy and reliability in detecting stress and emotional states. By leveraging the capabilities of AI, we intend to create a more accessible, cost-effective, and user-friendly solution that can be easily integrated into everyday life.

The primary objective of this study is to explore the potential of single-channel EEG headsets in emotion classification, focusing on their ability to capture the essential features of brainwave activity associated with different emotional states. This research will also examine the effectiveness of various AI algorithms in processing and interpreting EEG data, ultimately identifying the most suitable methods for real-time emotion recognition.

Furthermore, this study aims to contribute to the development of portable and affordable BCI systems that can provide real-time feedback on stress levels and emotional well-being. Such systems have the potential to revolutionize the field of mental health, offering individuals a practical tool for self-monitoring and stress management. By advancing the current state of BCI technology, we hope to pave the way for innovative applications that can significantly improve quality of life. This research endeavors to bridge the gap between complex multi-channel EEG systems and practical, everyday applications by developing a robust AI model for emotion classification using a single-channel EEG headset. The anticipated outcomes of this study include enhanced understanding of the relationship between brainwave patterns and emotional states, as well as the creation of a novel, market-ready solution for stress and emotion monitoring.

2. BACKGROUND

2.1. Brainwaves and EEG signals

Brainwaves are oscillating electrical voltages generated by nerve cells (also called neurons) of the brain. Since these cells are always active and rely on electrical impulses for communication with each other, brainwaves can be measured through an Electro-EncephaloGraphy (EEG) - a test that attaches small electrodes on the scalp to measure electrical activity in the brain. The data collected from an EEG through these electrodes is multidimensional and provides information about time, frequency, phase, and power of brainwaves. There are five types of brainwaves, including: alpha, beta, theta, delta, and gamma waves [1]. Nowadays, EEG is a valuable tool for testing hypotheses related to neurology and psychology because it can provide real-time information about brain health. In fact, EEG has been widely used in clinical diagnosis of epilepsy and other brain disorders.

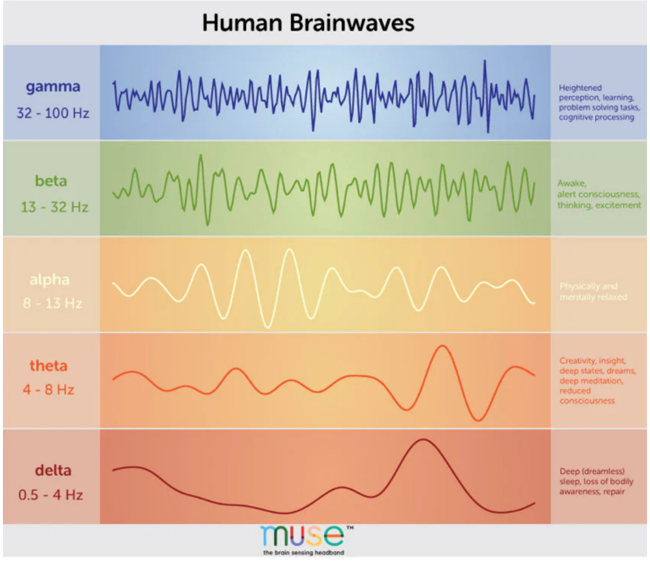


Figure 1: Human brain wave

Nerve cells generate electrical potentials through a process of so-called electrochemical polarization in the cell membrane. With the movement of ions across the channels of cell membranes, electrical currents are generated and creating potentials between the inside and outside of nerve cells. This process produces complex and diverse electrical signals that can represent brain activities [2]. With a large number of active nerve cells, the electrical signals (i.e., EEG brainwaves) on the scalp are recordable by electrodes placed on the scalp. Note that the electrical signals detected on the scalp by the electrodes are quite weak and require massive amplification as they pass through multiple layers such as skull, head skin, etc. For instance, when EEG data is used to classify different emotions, the electrodes Fp1, Fpz, and Fp2 need to be in the prefrontal areas of the head (behind which the prefrontal cortex is) because any electrical changes in these areas of the brain is strongly related to cognition and emotion [3]. Details about the placement of EEG electrodes are shown in Figure 3.

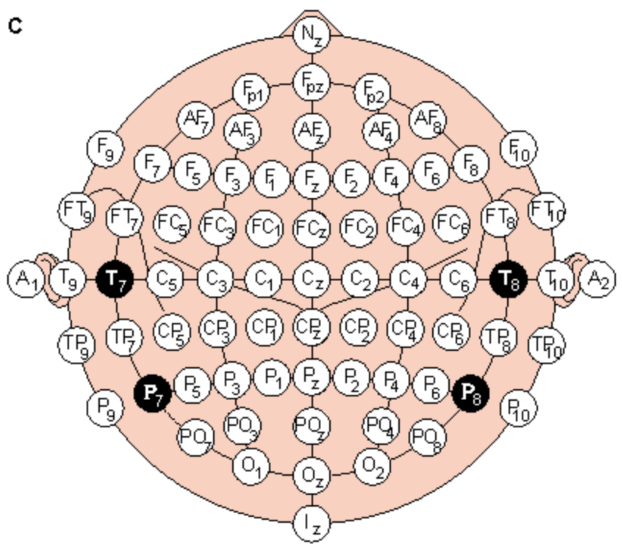


Figure 3: The 10-20 system of electrode placement [4]

2.2. Brain - Computer Interface

A Brain Computer Interface (BCI) aims for communication pathways between the computer or a device and the brain, based on neural activity generated by the brain. There are four main application areas of BCI systems: assisting disabled people, clinical monitoring for neurological diseases and sleeping disorders, neuroscience research on behavior, and human-machine interaction. There are some important properties [5] to take into account in order to design a BCI system:

- Noise and outliers: BCI features are noisy or contain outliers because EEG signals have a poor signal-to-noise ratio
- High dimensionality: in BCI systems, feature vectors are often of high dimensionality. Indeed, several features are generally extracted from several channels and from several time segments before being concatenated into a single feature vector
- Time information: BCI features should contain time information as brain activity patterns are generally related to specific time variations of EEG
- Non-stationarity: BCI features are non-stationary since EEG signals may rapidly vary over time and more specifically over sessions
- Small training sets: The training sets are relatively small, since the training process is time-consuming and demanding for the subjects (not have to be true for clinical use). The basic steps for building a BCI system for human- machine interaction are 1) acquire the EEG signal, 2) process and classify the EEG signal, 3) use the classified signal to control the interface, and 4) present feedback to the user.

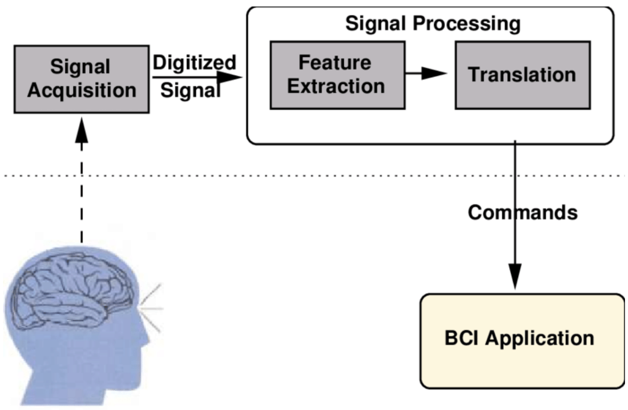


Figure 4: A BCI system design

2.3. The mechanism of brainwave spectrum changes during emotional stress classification

The definition of stress has been a topic of ongoing debate since it was first proposed, and no universally accepted definition exists [6]. In this project, we follow the definition given by World Health Organizationn (WHO)

[7]: Stress can be defined as a state of worry or mental tension caused by a difficult situation. The stress response can be measured from perceptual and physical human responses [6]. Apart from the perceptual responses, when the body is under the stress circumstances, it releases stress-related hormones to cause hormonal changes in the body. Hence, the physical responses can be used as indicators to measure or diagnose the stress. It has been shown that stress can be assessed from physiological variables including EEG [8], blood pressure [9], heart rate variability [10], and various human indexes.

As hormonal changes have a direct impact on brain activity, the electroencephalogram (EEG) index is commonly employed in various studies to investigate this relationship. EEG signals can be used to identify human stress levels, as research has established a significant correlation between psychological stress levels and EEG power. For instance, a study [11] demonstrated that stress is positively correlated with beta EEG power in the anterior temporal lobe. Additionally, research suggests that the left hemisphere is involved in processing positive emotions, while the right hemisphere plays a more prominent role in processing negative emotions and withdrawal behaviors [12]. These hemispheric differences are represented by a model of emotional processing in which the frontal cortex plays a crucial role [13].

Furthermore, research [14] has revealed that positive moods or reactions are associated with relatively greater left prefrontal activity (LFA), while negative moods or reactions are linked with relatively greater right prefrontal activity (RFA). Similar effects have been reported for stress-related emotions, with preferential right hemispheric activation in the frontopolar region [15].

Various techniques have been employed to induce and study stress levels in laboratory settings, where EEG signals are recorded. For example, the study [6] used the Stroop color-word test as a stressor to induce four levels of stress. In another study [16], the Trier Social Stress Test (TSST) was used to create a stressful environment to explore the causes of false recognition. Additionally, IQ-based questions have been used to induce different stress states in various experiments [17, 18]. An experiment [19] utilized emotion classification methods to assess stress levels. The study by Aftanas et al. [20] showed significant changes in EEG signals collected from participants watching images with different levels of arousal. Hosseini et al. [21] used visual images to induce two emotional stress states (negative and positive). Kim et al. [22] developed a stress state classification system using EEG data collected from music and story stimuli.

3. RELATED WORK

To prepare theoretical bases for this project, we reached out for existing studies and reviews regarding usage of EEG signals combined with machine learning models to identify and quantify stress. Results summarized from these documents allowed us to have a more detailed view of which models should be expected to best capture our problem. Some noticeable work we have found are as follows:

Saeed et al. [28] conducted an experiment on 33 subjects to identify rest and stressed mental states using neural oscillatory features and alpha – beta asymmetry features extracted from five frequency bands: delta, theta, alpha, beta, and gamma. They then used several Machine Learning models for classification, including SVM, Naive Bayes, KNN, Logistic Regression, and MLP, all of which resulted in above 80% accuracy (with the exception of KNN which only had a low 65.96% accuracy).

Khosrowabadi et al. [24] also experimented to differentiate rest and stressed states on 26 subjects, using the entirety of frequencies from 2 to 32 Hz for feature extraction: Higuchi's fractal dimension, Gaussian mixtures of EEG spectrogram, and MSCE. Passing the features to KNN and SVM, they acquired a 90% accuracy on both models and noticed the significance of MSCE as a validation criterion between different subjects to classify the EEG.

Masood et al. [31] utilized a Convolutional Neural Network (CNN) to learn from the Power Spectral Density (PSD) of four frequency bands (theta, delta, alpha, beta), and were able to classify relaxed and stressed subjects with 87.5% accuracy. They also noted that adding more measurements such as Heart Rate Monitor (HRM) and Respiration Rate (RESP) will increase stress detection accuracy.

With knowledge provided by these findings, we centered our project around some main methods of data acquisition and model selection, along with mathematical strategies to preprocess the EEG signals.

4. METHODOLOGY

4.1. Data Acquisition

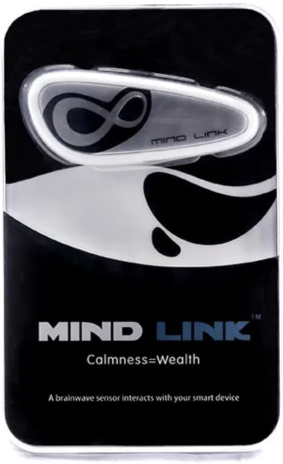


Figure 5: Mindlink

In our experiment, EEG data is collected from 5 subjects. All subjects have no history of mental diseases and head injuries. A wireless EEG device Mind-link is used to record EEG signals. The device is designed to have three electrodes (Fp1, Fpz, Fp2) and collect data from **single channel** Fp1 placed on the prefontal lobe following the international 10-20 system as presented before, sampled at 512 Hz with bandwidth from 0.5 Hz to 40 Hz.

Figure 6: Electrodes of the wearable headset

The data acquired from the MindLink will be transmitted to the ESP32 via Bluetooth signal through the HC05 module in the form of bits. The data bits will be decoded and converted into data values ranging from -255 to 255.

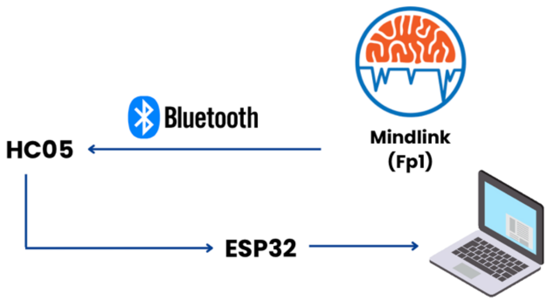


Figure: EEG signal flow

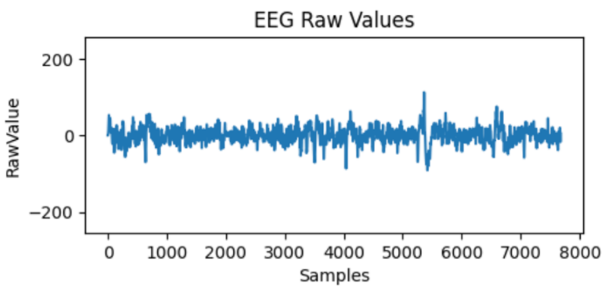


Figure 7: Raw EEG signal

Motivated by the data acquisition processes in [21], [22], in this project, we combine both visual and auditory stimuli to induce varying stress levels in a participant. Each participant experiences a 6-minute process that can be described as follows: 1 minute of relaxation, 2 minutes watching 'positive mood' music videos (MV), another 1-minute relaxation, and 2 minutes watching 'stress-inducing' MV. So the total length of our dataset is 30 minutes - 921,600 data points. The relaxation period before experiencing any emotional stress is helpful for calming and refreshing the mental state. Figure 4 describes the data acquisition process. In this experiment, a 'positive mood' MV is a music video that has upbeat and exciting audio accompanied by humorous scenes. On the other hand, a 'stress' MV contains distressing beats and lyrics, as well as negative scenes depicting death, suicide, or self-destructive behavior. The list of MVs used in this project is shown in Table 1. Since the following list MVs consists of well-known and familiar MVs among the participants, they can comprehend the content and clearly distinguish between the 'stress' and 'goodmood' MVs. This is crucial because if unfamiliar MVs were used, their content might cause confusion and easily affect the data labeling process. Moreover, these MVs possess a relatively distinct and consistent level of emotional impact, enabling participants to select the MV based on their personal preferences.

Figure 8: Data acquisition diagram

Music video	Emotional stress state
Billie Eilish, Khalid - lovely	Stress (Distressing beat and lyrics, sad scenes)
Joji - Glimpse of Us	Stress (Self destruction scenes)
Lewis Capaldi - Wish You The Best	Stress Scenes of parting, loss of loved ones
Mark Ronson, Bruno Mars - Uptown Funk	Goodmood (Energetic, vibrant funk choreography; Lively, upbeat music)
Maroon 5 - Sugar	Good mood (Happy, marriage scenes; lively and upbeat melody)
Pharrell Williams - Happy	Good mood (Joyful dancing scenes; feel-good, happy beat and lyric)

Table 1: List of Music Videos

4.2. Feature Extraction

4.2.1. Time Domain

A time-domain analysis is a time-based analysis of physical signals, mathematical functions, or time series of economic or environmental data. In the case of discrete-time or continuous-time, the signal or function's value is also understood for all real numbers at multiple separate instances in the time domain. The transitory response of a system to be investigated is provided through time-domain analysis, which allows for a better understanding of the flow of both mechanical and electrical energies. Wave propagation, system structural changes, and electric potential generated by external excitations are all examples of this. As a result, the time-domain technique serves as a link between traditional spectral analysis and physical time interpretation. Furthermore, time-domain methods allow for on-the-fly measurement of basic signal parameters using a time-based calculation, which requires less complex equipment than traditional frequency analysis.

We used time_series_features, Hjorth_features and fractal_features:

Time_series_features :

Variance:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N - 1}$$

RMS:

$$RMS = \sqrt{\frac{\sum_{i=1}^{N}(x_i)^2}{N}}$$

Peak-to-Peak:

$$Ptp = |x_{max} - x_{min}|$$

Hjorth_features :

$$Mobility = \sqrt{\frac{var(dx(t))}{var(x(t))}}$$

$$Complexity = \sqrt{\frac{var(dx^2(t))}{var(x(t))}}$$

Fractal_features:

Katz_fractal

$$FD = \frac{\log N}{\log L}$$

Higuchi_fractal

$$FD = \frac{\log lk}{\log k}$$

4.2.2. Frequency Domain

Fast Fourier Transform (FFT) is a technique that transforms a signal from time domain to frequency domain. Using FFT, we were able to transform sequences of EEG signals, recorded in time domain, into frequency domain for desirable feature extractions.

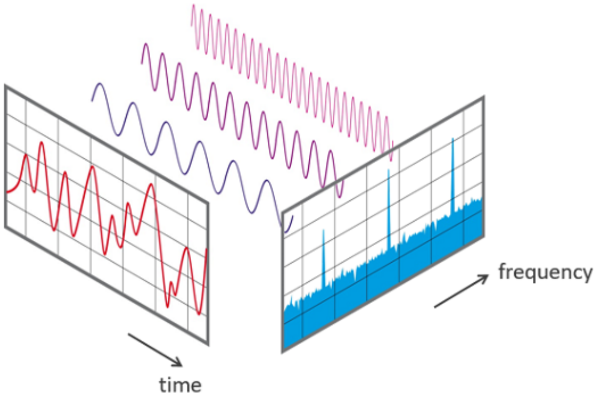


Figure 9: Demonstration of FFT

The frequency spectrum allowed us to analyse how power density is distributed on different frequency modes. Combined with theoretical bases on brainwaves and the changing mechanism of brainwave frequency spectrum presented in later parts of this report, we can identify the emotional state of a person to be stressed or relaxed.

After transforming the signals to frequency domain, four frequency bands are extracted and their powers are calculated as below, with the range of the bands being: delta ($0.5 \leq f \leq 4$), theta ($4 \leq f \leq 8$), alpha ($8 \leq f \leq 13$), and beta ($13 \leq f \leq 30$). Ratios between powers of these frequency bands are also taken into consideration to determine one's emotional state.

$$E_\delta = \sum_{freq = 0.5}^{4} P_{freq} \tag{1}$$

$$E_\theta = \sum_{freq = 4}^{8} P_{freq} \tag{2}$$

$$E_a = \sum_{freq = 8}^{13} P_{freq} \tag{3}$$

$$E_\beta = \sum_{freq = 13}^{30} P_{freq} \tag{4}$$

$$ABR = \frac{E_a}{E_\beta} \tag{5}$$

$$TBR = \frac{E_\theta}{E_\beta} \tag{6}$$

$$DBR = \frac{E_\delta}{E_\beta} \tag{7}$$

$$TAR = \frac{E_\theta}{E_\alpha} \tag{8}$$

$$DAR = \frac{E_\delta}{E_\alpha} \tag{9}$$

$$DTABR = \frac{E_\delta + E_\theta}{E_\alpha + E_\beta} \tag{10}$$

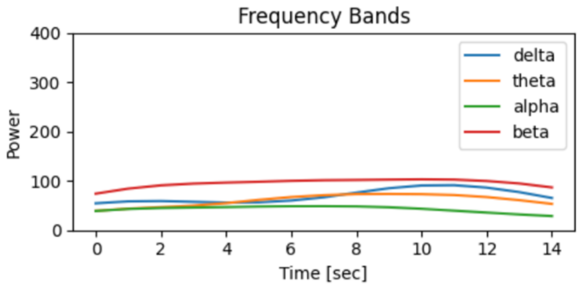


Figure 10: Power of 4 frequency bands

4.2.3. Time-frequency Domain

Time-frequency analysis techniques are a class of algorithms that enable extracting more information from a signal in the time domain, which typically contains only time and amplitude data. By applying time-frequency analysis to time-domain signals, both temporal and frequency characteristics can be obtained. The resulting data is represented as multiple time-varying signals at different frequencies, allowing the computation of time-frequency texture descriptors from single-channel EEG signals. Short-Time Fourier Transform (STFT) and Wavelet transforms are among the most widely used time-frequency analysis methods.

EEG signals exhibit non-stationary behavior, meaning their spectral content changes over time. While the traditional Fourier Transform (FFT) can reveal the frequencies present in a signal, it does not provide information about when these frequencies occur. STFT involves computing the FFT over short, overlapping segments of the signal, known as windows, allows the identification of the time intervals during which specific frequencies are present in the signal. The STFT is computed using a window function w centered at time n, given as:

$$I_X(n,k) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{\frac{-j2\pi km}{N}} \quad (11)$$

where x[m] is the signal to be analyzed and N is the frequency sampling factor. The resulting STFT matrix I is represented as a matrix with time and frequency $\omega = 2\pi k/N$ information. [32]

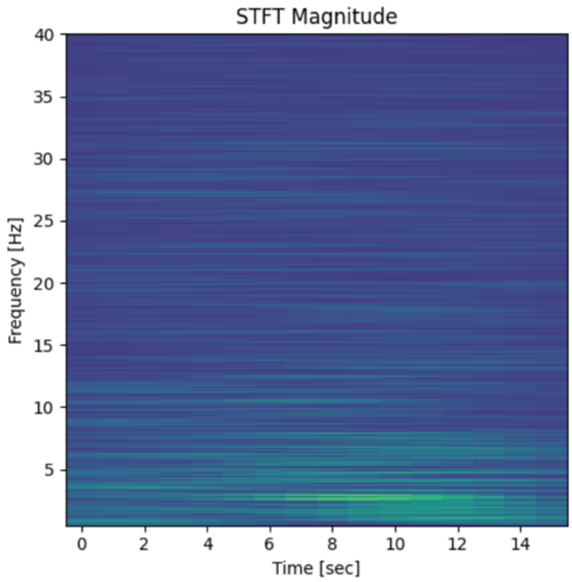


Figure 11: STFT spectrogram from EEG signal

The Gray-Level Co-occurrence Matrix (GLCM) is a powerful statistical tool widely used in texture analysis and feature extraction from gray-level images. It is a matrix that represents the frequency of occurrences of different combinations of pixel brightness values (gray levels) in an image. The GLCM is calculated by considering the spatial relationship between pixels, typically in a specific direction (0°, 45°, 90°, or 135°) and distance (offset).

The GLCM can be mathematically represented as a square matrix, where the number of rows and columns corresponds to the number of gray levels in the image. The element (i, j) of the GLCM represents the frequency of occurrences of pixels with gray level i and j, separated by the specified offset and direction.

$$P_{\delta,\theta}(i,j) = \sum_{n=1}^{N}\sum_{m=1}^{K} \begin{cases} 1 & \text{if } I(n,m) = i \text{ and } I(n+_{\theta}, m+\delta) = j \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where i, j = 0, 1,…, L-1; L is the number of gray scales; N and K are the sizes of the STFT matrix. [32]

Once the GLCM is calculated, various statistical features can be derived from it, such as contrast, correlation, energy, homogeneity, and entropy which formulas are represented in the Figure 9 [23]

No.	# of features	Feature name	Equation
1	1 feature	Contrast	$\sum_{i,j} \|i-j\|^2 p(i,j)$
2	1 feature	Correlation	$\sum_{i,j} \frac{(i-\mu_i)(i-\mu_j)p(i,j)}{\sigma_i \sigma_j}$
3	1 feature	Energy	$\sum_{i,j} (p(i,j))^2$
4	1 feature	Homogeneity	$\sum_{i,j} \frac{p(i,j)}{1+\|i-j\|}$
5	1 feature	Entropy	$\sum_{i,j} p(i,j) log_2 p(i,j)$

Figure 12: Features of GLCM

Where μ_i and μ_j are the mean values according to i and j references respectively, respect to the 1 and 2:

$$\mu_i = \sum_{i,j} p(i,j) * i \quad (13)$$

$$\mu_j = \sum_{i,j} p(i,j) * j \quad (14)$$

and σ_i, σ_j are the standard deviation of values for i and j references respectively.

Over the years, the Wavelet Transform (WT) has been frequently used in time-frequency analysis as an alternative to the traditional Short-time Fourier Transform (STFT) technique, which is theoretically based on the Fourier Transform (FT). Therefore, it is natural that the idea behind this technique, as well as its advantages, are better understood through comparison with FT and STFT.

4.2.3.1. Limitations of the Fourier Transform

The Fourier Transform is a tool that breaks down a waveform (a signal) into a sum of sinusoidal signals (periodic functions based on the sine or cosine function). Given a signal $x(t)$, its Fourier Transform is given by the formula:

$$F(k) = \int_{-\infty}^{+\infty} e^{-2\pi ikt} x(t)dt$$

Thanks to the Euler formula, we can rewrite the formula as:

$$F(k) = \int_{-\infty}^{+\infty} x(t)(cos(2kt) - isin(2kt))dt$$

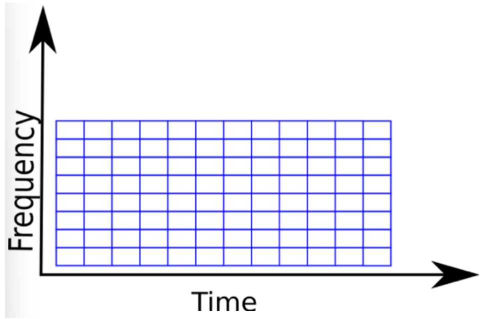
Where k is some frequency. As a result, $F(k)$ gives how much power $x(t)$ contains at the frequency k.

As can be seen, the building blocks of FT (the complex exponential) oscillate over all of time. For this reason, it is difficult for FT to represent signals that are localized in time. By taking an FT of a time signal, time information is lost in return for frequency information.

On the other hand, EEG signals have a very high temporal resolution, on the order of milliseconds. FT simply can't exploit this fact.

4.2.3.2. Limitations of STFT

To capture both time and frequency information, the STFT is developed. In the STFT, we perform a series of windowing and FT operations. At each time shift t, we apply a window $w(t' - t)$ centered around t', and then calculate the Fourier Transform of $f(t')w(t' - t)$. This way, we can observe changes in the power spectrum over time.



For a function $f(t)$, the STFT is defined as follow:

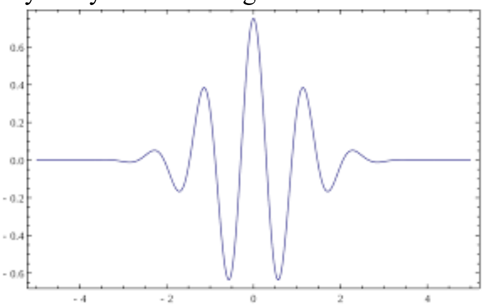
$$\hat{f}(t,u) = \int_{-\infty}^{\infty} f(t')w(t'-t)e^{-i2\pi t'u}dt'$$

The choice of windows presents an important trade-off, which is also known as the uncertainty principle. If the windows are wide, one can get good resolution in frequency but end up with poor resolution in time. In the case of FT, we lost all information about time. Conversely, if windows are narrow, then the signal is better resolved in time, but we lose resolution in frequency. Since the windows are of a fixed size with a predefined time shift t, STFT gives a fixed temporal resolution for all frequencies (low and high), which can't adapt to signals where the frequency fluctuates a lot.

Therefore, it would obviously be better if we could decompose our signal in a way that high-frequency components are analyzed with high spatial resolution (narrow windows) since they vary rapidly in time, and low-frequency ones are analyzed with low temporal resolution (wide windows). This better way is the Wavelet Transform.

4.2.3.3. Continuous Wavelet Transform (CWT)

It is worth noting that STFT still uses sinusoids that are active for the entire time window (shown in the complex component). Wavelets, on the other hand, are families of signals that are sort of like a sinusoid, but only for a brief period of time and zero elsewhere. One of the most common wavelet families used in the Time-frequency analysis of EEG signals is the Morlet wavelet.



Note that a wavelet is actually a complex quantity as it is obtained by multiplying a complex sinusoid with a Gaussian. The figure above only shows the real part for the sake of visualization.

The Continuous Wavelet Transform of a signal $x(t)$ is given by:

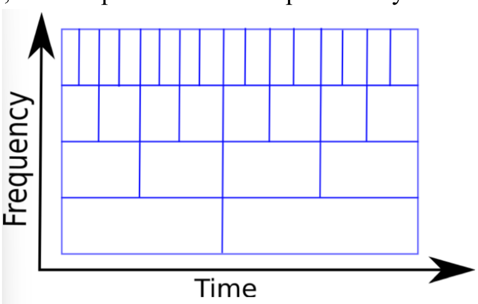
Continuous Wavelet Transform (CWT)

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^{*} \frac{(t-b)}{a} dt$$

a: Scale (or dilation) parameter
b: Location of wavelet
Ψ: Wavelet function
x: Signal

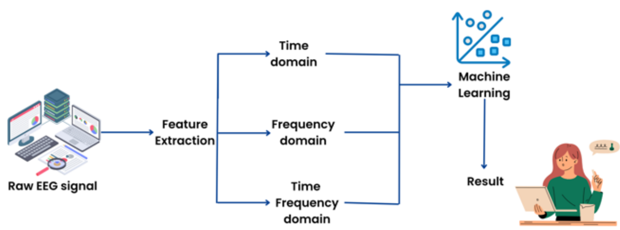
Wavelet transformation essentially involves convolving the complex wavelet with the EEG signal and moving it along the time axis (known as a translation) by b units and doing this with wavelets of the same family but with varying frequencies, representing through the scale a. There are different functions used to convert scales to frequencies, but they all indicate an inversely proportional relationship between scales and frequencies. So, higher scales are associated with lower frequency and lower scales are associated with higher frequency.

As a result, unlike the STFT which has a window size throughout, with wavelets we can analyze a signal at different frequencies with different resolutions. $T(a,b)$ is the Wavelet coefficient, it tells you how much of the "daughter" wavelet function is present in the signal at the scale a, and the position in time specified by b.



As seen from the figure above, wavelets can provide good frequency resolution and relatively poor temporal resolution at low frequencies, while good temporal resolution and relatively poor frequency resolution at high frequencies.

4.3. Machine Learning Algorithms



Solution diagram for emotional stress classification

4.3.1. Support Vector Machine (SVM)

SVM is a learning algorithm used in supervised learning and mostly for classification problems by looking for a hyperplane that can separate well the classes of a given dataset. One advantage of SVM is that it can work well with very high dimensional problems, which is the case for this project where multiple features are extracted from the EEG signals. Khosrowabadi et al. [24] and Sani et al. [25] applied SVM to quantify two levels of stress and achieved accuracy levels of 75% and 90%, respectively, while according to Hou et al. [26], accuracy of SVM tend to decline as the number of stress levels increase.

Two hyperparameters of the model were thoroughly examined: the penalty constant C showing how much attention the model should pay to noises or outliers, and the kernel function used for transforming data to find a feasible hyperplane. By utilizing Grid Search cross-validation to look for optimal hyperparameters, we selected C to be 100 and the radial basis function (RBF) as the kernel function, which gave the most promising results with **frequency domain** features. For **time-frequency domain** features we have to decrease the C to 50 as now we have more features to consider. Additionally, we define the gamma parameter equals to 1 to ensure that each training example has a certain level of influence that is moderate, neither too large nor too small.

4.3.2. Gradient Boosting

We examined two boosting algorithms for their viability in this problem: AdaBoost and Gradient Boosting. However, AdaBoost performed poorly due to its nature of being sensitive to noises: mislabeled datapoints are weighted more and more as they go through many weak learners, causing overfitting in the model. Therefore, only Gradient Boosting was considered in this project.

Here, Gradient Boosting uses Decision Trees of low terminating depth as the weak learners. Each new estimator is trained to minimize a loss function using gradient descent, such as mean squared error (MSE) or cross-entropy loss, given the output of the previous estimator as its input. In contrast to AdaBoost, the weights of the training instances are not tweaked. Instead, each estimator is trained using the residual errors of the predecessor as labels, and the gradients used to update the weights are less sensitive to outliers; therefore, Gradient Boosting is more robust.

4.3.3. Multilayer Perceptron (MLP)

MLP is a feed-forward artificial neural network consisting of fully connected neurons with a nonlinear activation function, notable for being able to distinguish data that is not linearly separable. Arsalan et al. [28] found that MLP outperforms both SVM and NB classifiers and gives the highest accuracy for quantifying mental stress at both two and three levels. Meanwhile, Saeed et al. [27] reported that integrating alpha, beta and gamma features with MLP provides the highest accuracy (85.13%) compared to that achieved using a single feature.

However, the main drawback of MLP is potential over-fitting due to excessive or insufficient neurons [28]. Therefore, careful selection of the neural network's structure is required, including the number of neural layers and perceptrons per layer, and the activation function to work well with the extracted features from EEG signals.

4.3.4. K-Nearest Neighbour (K-NN)

K-Nearest Neighbour (K-NN) classifies datapoints based on the classes of its neighbours. The new datapoints are classified by the model using the similarity measure of the previously stored data, meaning that the mechanism of K-NN depends on estimating the "distance" between neighbours and choosing the K closest neighbors. Thus, two of the critical factors to be identified are the optimal value of K and neighbors distance D [27, 29].

K-NN has a non-parametric learning technique, meaning that it has a much lower computational complexity in quantifying mental stress, especially with small-sized data [29, 30]. We apply this algorithm with both **time domain** and **time-frequency domain** features. For time domain, the result obtained is good enough, with non-parametric K-NN. On the other hand with time-frequency domain features, we selected L1 distance and K to be 7 by utilizing Grid Search cross-validation to have the best result. The L1 distance here seem to be reasonable as time-frequency domain features is non-linear. The 'weights' parameter is set to 'distance' to ensure that the nearer neighbor has more effect on the prediction than the further neighbor.

4.3.5. Decision Tree

Decision Trees are tree-like models where internal nodes represent feature tests, branches represent decision rules, and leaf nodes represent the final predicted classes or values. Decision Trees work by recursively partitioning the input space into smaller regions based on the most discriminating features, until a desired stopping criterion is met. During prediction, a new data instance is traversed through the tree from the root node to a leaf node, following the path determined by the feature tests at each internal node. [33]

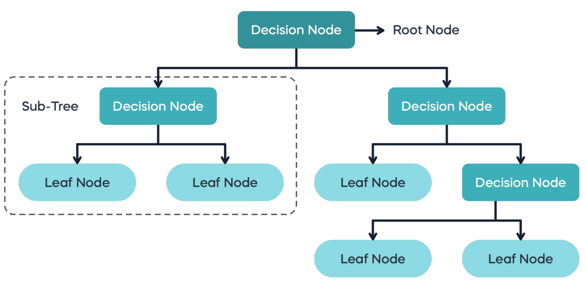
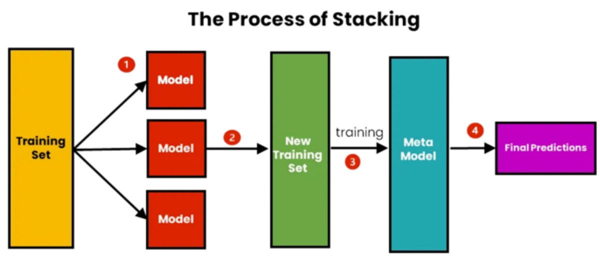


Figure 13 : Decision Tree algorithm [34]

In this project we only apply the Decision Tree with STFT and GLCM features. By applying Grid Search cross-validation, the most promising result is obtained with a set of parameters as follow: {'criterion': 'entropy', 'max_depth': 7, 'min_samples_split': 5}, other parameter is defaultly set. It is a tradeoff when using 'entropy' (information gain) compare to 'gini' when measuring the quality of a split in the tree. Entropy is more computationally expensive but preferable for small datasets. This is also suitable for our multilabel classification task. Other parameters are set to deteriorate the overfitting problem. During the training process, we also try to apply post-pruning technique to gain a better result but there is no significant change.

4.3.6. Ensemble Learning

Ensemble learning is a powerful technique that combines the predictions of multiple machine learning models to improve overall performance. In this study, we employ stacking, an ensemble method that leverages the strengths of various base learners and a meta-learner to achieve superior accuracy.



Three base learners were selected for their complementary characteristics:

- **Support Vector Machine (SVM):** Effective in high-dimensional spaces, SVM models complex decision boundaries, particularly with a radial basis function (RBF) kernel. Probabilistic outputs were enabled for ensemble integration.
- **Random Forest (RF):** This ensemble of decision trees mitigates overfitting, offering robustness and generalization capabilities.
- **Gradient Boosting (GB):** Sequential error correction and optimization of loss functions make GB adept at handling complex interactions and non-linearities.

Each model is fine-tuned using GridSearchCV to optimize their hyperparameters as follows:

- **Support Vector Machine (SVM)**
  - Hyperparameters:
    - C: Controls the trade-off between achieving a low training error and a low testing error [0.1, 1, 10, 100]
    - gamma: Defines how far the influence of a single training example reaches ['scale', 'auto']

        - kernel: Specifies the kernel type to be used in the algorithm ['rbf', 'linear']
    - Best Parameters (from GridSearchCV):
        - C = 0.1
        - gamma = scale
        - kernel = rbf
- **Random Forest (RF)**
    - Hyperparameters:
        - n_estimators: The number of trees in the forest [100, 200, 300]
        - max_depth: The maximum depth of the tree [10, 20, 30]
        - min_samples_split: The minimum number of samples required to split an internal node [2, 5, 10]
    - Best Parameters (from GridSearchCV):
        - n_estimators = 100
        - max_depth = 10
        - min_samples_split = 10
- **Gradient Boosting (GB)**
    - Hyperparameters:
        - n_estimators: The number of boosting stages to be run [100, 200, 300]
        - learning_rate: Shrinks the contribution of each tree [0.01, 0.1, 0.2]
        - max_depth: The maximum depth of the individual regression estimators [3, 4, 5]
    - Best Parameters (from GridSearchCV):
        - n_estimators = 100
        - learning_rate = 0.01
        - max_depth = 3

The meta-model used is Logistic Regression with L2 regularization (Ridge). This choice is motivated by the need to balance model complexity and interpretability while preventing overfitting. L2 regularization adds a penalty term to the loss function, proportional to the sum of the squared coefficients.

The stacking ensemble model is constructed using the best estimators from the GridSearchCV for each base model. The training process involves:

- Training Base Models: Each base model is trained on the training data.
- Training Meta-Model: The predictions from the base models are used as features to train the meta-model.

To ensure the robustness of the model, stratified K-fold cross-validation is performed. The data is split into 10 folds, ensuring each fold has a representative distribution of the classes. The model is trained and evaluated on each fold, and the average accuracy across the folds is reported.

The performance of the stacking classifier was assessed using accuracy, precision, recall, and F1-score. It achieves high precision, recall, and F1-scores across all classes. Additionally, confusion matrices were generated to visualize the model's performance across different classes. These metrics provide a comprehensive evaluation of the model's effectiveness in stress detection from EEG signals. The confusion matrix reveals that the "Calm" class is most accurately identified, while there is some confusion between the "Stressed" and "Good mood" classes.

Overall, the ensemble learning approach with stacking and Grid Search provided a robust framework for accurately detecting stress levels from EEG signals, leveraging the strengths of multiple machine learning models to achieve high predictive performance.

To enhance the model's performance, several strategies can be pursued. Initially, employing more stringent regularization techniques could mitigate the potential overfitting indicated by the high training accuracy. This could involve methods such as dropout, or early stopping. Subsequently, augmenting the dataset through techniques like time shifting, noise addition, or frequency filtering could improve the model's ability to generalize to unseen data. The increased size and diversity of the dataset would expose the model to a wider range of variations, leading to improved robustness. Additionally, refining the feature extraction process could yield more discriminative representations of the EEG data. This could involve exploring alternative feature types (e.g., time-frequency features, connectivity measures) or applying more sophisticated feature selection algorithms. By enhancing the class separability in the feature space, classification performance could be improved.

In conclusion, while the current stacking ensemble model demonstrates promising results, there remains potential for further optimization. Systematic exploration of hyperparameter tuning and consideration of alternative models or ensemble strategies could lead to incremental improvements in accuracy and robustness.

4.3.7. Simple Neural Network

Based on our analysis of the data, we observed that the mean and variance change over time. Consequently, we concluded that the time series is not stationary. Since stationarity is a prerequisite for many linear models, this indicates that the problem at hand is not linear. As a result, we decided that a simple neural network would be more suitable for this task.

Given the small size of the dataset, we utilized the repeat(10) method to augment the data and provide more samples for training. However, during training, we noticed that the validation accuracy consistently decreased compared to the training accuracy after each epoch. This behavior suggests that the model was overfitting to the training data. To mitigate this, we implemented regularization techniques, specifically L2 regularization. Regularization helps to prevent the model from becoming too complex and thus improves its generalization ability to unseen data.

Moreover, we opted to use integer encoding rather than one-hot encoding for the target variables. The rationale behind this choice is based on the semantic relationship between the categories. For instance, in the context of emotion classification, the label "sad" is semantically closer to "calm" than to "happy".

Based on further experiments, we determined that the optimal window size for this problem is 30 seconds. This finding suggests that the emotion at any given moment is influenced by the preceding half-minute of data.

4.3.8. Convolutional Neural Network (CNN)

4.3.8.1. Input generation

Our EEG signals are sampled at 512 Hz and stored in different text files, which are categorized by the participants in the recording process and the states they were in.

First, we extract smaller EEG signals using a sliding window of size 5×512 overlapped by 4×512. This means that we will try to categorize EEG signals recorded in 5 seconds from the start with a shift of 1 second. Afterwards, the signals are normalized.

By applying the Continuous Wavelet Transform through the function `pywt.cwt(signal, scales = range(1, 256), wavelet = 'morl')`, we get the Wavelet coefficients with the shape `(len(scales), len(signal))`. These coefficients will be used as the input for the CNN.

CNN is another type of commonly used neural network. It would be sensible to first talk about a regular neural network first.

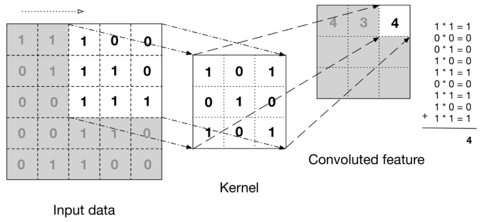
4.3.8.2. Neural Network

Neural networks are modified after our brains. Just like how we can distinguish between a red and blue pen by receiving information from our eyes and then processing it through several neurons connected, there are "nodes" that form the layers in a neural network. The nodes contain values in them which are called "activation".

A weight (a real number) is assigned to each link between every two nodes in adjacent layers. The network takes all of the training data in the input layer and passes them through the hidden layers, generating the "activation" at each node by matrix multiplication. Finally, it returns a value in the output layer.

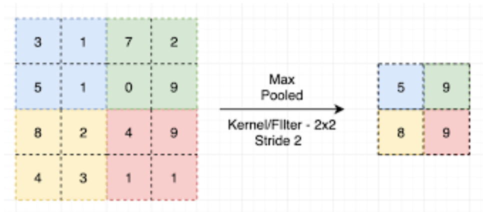
4.3.8.3. Convolutional Neural Network

A big difference between a CNN and a regular neural network is that CNNs use Convolution and Pooling layers to handle the math. A convolution is used instead of matrix multiplication in at least one layer of the CNN. The image below shows what a convolution is.



Basically, we apply a filter to the input image and shift it to get the convolved feature by the Hadamard product. This convolved feature is passed on to the next Pooling layer. One can understand the convolved feature as a container of the most relevant information we wish to detect. What makes CNNs so special is that they are able to tune the filters as the training happens, this removes the need for hand-created filters.

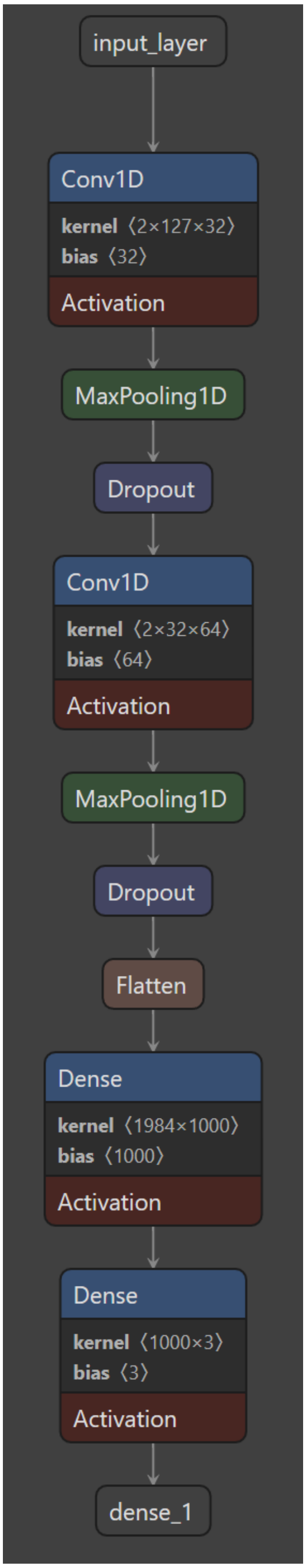
The Pooling layer is responsible for the reduction of the Convolved feature's dimension. This is to decrease the computational power required to process the data. In our CNN, only the Max pooling layers are used. What they do is slide through the convoluted feature and find the maximum values, performing as a noise suppressant (only the most noticeable features are selected). Here is an example of how a Max Pooling layer works.



This is what gets passed to each successive Conv – MaxPool layers until we get to the final classification layer. To reduce overfitting, `Dropout(rate = 0.5)` layers can be implemented. Since overfitting is when the model learns the statistical noise in the training data, dropping out some outputs of a layer makes the training process noisy. The `rate` hyperparameter represents the probability that outputs of the implemented layer are dropped out or retained randomly in the forward propagation. Since we are applying this for hidden Convolutional layers, a common value of 0.5 is chosen. L2 regularization is also applied in the first Dense layer.

Below is the structure of the CNN model used for this problem:

(figure below)



5. RESULTS

Models	Feature extraction	Test accuracy	Train accuracy
MLP	Frequency domain	0.867	0.95375
SVM	Frequency domain	0.86111	0.9
	Time-Frequency domain	0.95454	1
Gradient Boosting	Frequency domain	0.905	1
StackModel	Frequency domain	0.9125	0.99688
K Nearest Neighbors	Time domain	0.96363	0.97648
	Time-Frequency domain	0.92614	0.95092
Decision Tree	Time-Frequency domain	0.88068	0.97866
Random Forest	Time-Frequency domain	0.91761	1
Convolutional Neural Network (CNN)	Time-Frequency domain	0.68182	0.70237
Artificial Neural Network	Frequency domain	0.99	1.00

Table 2: Accuracy of machine learning models

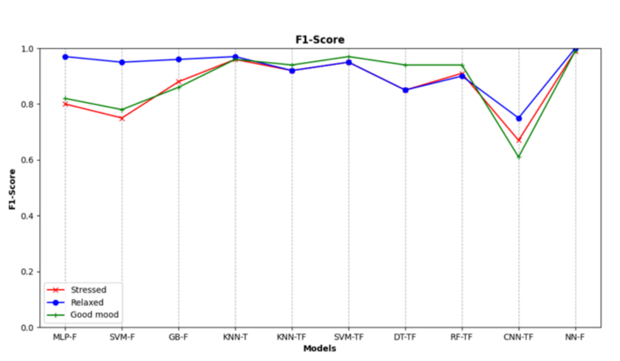


Figure 14: F1 score of all models

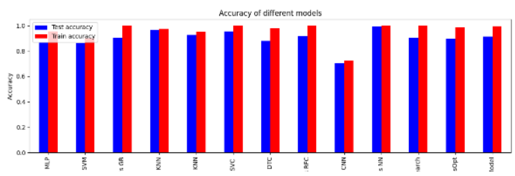


Figure 15: Accuracy of all models

Out of the three types of feature extraction, the frequency domain seems to be the most informative as the input of many models with great test and train accuracy, including the Neural Network model with 99% and 100% test and train accuracy respectively.

For this dataset, time-frequency domain extraction is not suitable due to its computationally expensive nature and the small dataset size. This is evident in the application of CWT and CNN in classifying the EEG signals, this model performed worst with only 68% and 70% test and train accuracy respectively.

One obvious weakness of CNN is that it requires a lot of data. Since our data is self-generated, along with the preprocessing phase to extract time-frequency data in the form of Wavelet coefficients, the training data for our CNN model doesn't exceed 2000. Due to its computationally expensive nature, the "scales" parameter used to generate daughter wavelets with different frequencies is set to only receive 127 values. In the process of building this model, we realize that the more scale values are, the more information is acquired about our signals, which in turn will help our model learn better.

Other models prove to be highly effective in this classification task with more than 85% test accuracy.

6. DISCUSSION AND FUTURE WORK

Throughout this project, we have demonstrated the steps required to identify whether a person is mentally stressed: measuring brainwaves using EEG, filtering and extracting desired features, and using a Machine Learning model to learn and predict the mental state from the extracted data. These steps form a pipeline-like system for our problem, and they can all be improved with further investigation to yield better results.

Gathering a larger dataset allows the models to better generalize underlying patterns of the EEG signals. Different persons might have different power levels of brainwaves, so recording from more subjects can further assure the generalization of the models. Experiment conditions should also be considered, from stress stimuli to potential external noises, to ensure a well-controlled environment for the recording process, resulting in better data quality.

During the training process, more advanced models may produce better outcomes, especially Deep Learning models such as Long Short-Term Memory (LSTM) networks or Recurrent Neural Networks (RNNs). Advanced machine learning models, such as Transformer models, should be explored for their potential to capture temporal dependencies. Utilizing transfer learning techniques can also improve performance with limited datasets.

Applications of this topic in real life are also worth exploring, such as assessing stress for proper treatment in stressful environments like academic settings, long-distance driving, or office workplaces. Early detection of stress can help people cope better with it, ensuring good healthcare in daily life.

Future research in EEG-based mental stress detection should address several key areas to enhance efficacy and practicality. One primary limitation is the small dataset size. Collecting data from a larger and more diverse pool of participants will improve the model's generalizability. Including varied demographics can provide a comprehensive understanding of mental stress manifestations.

Enhancing EEG signal processing techniques, such as Independent Component Analysis (ICA), and developing real-time signal processing algorithms are crucial for practical applications. Integrating EEG data with other physiological signals like heart rate variability (HRV) and incorporating contextual data can improve the robustness and accuracy of stress detection systems.

Developing user-friendly, wearable EEG devices that are lightweight, non-invasive, and aesthetically pleasing will enhance user adoption. Ensuring seamless integration with mobile devices and cloud platforms is also essential for real-time monitoring and feedback.

Conducting extensive field studies in real-world scenarios will help validate the models and systems. Developing personalized stress management strategies, such as biofeedback and mindfulness training, based on real-time stress detection is another important area for future research.

Ensuring the privacy and security of EEG data through robust encryption and anonymization techniques is paramount. Addressing ethical concerns related to brainwave monitoring technology is also crucial.

By focusing on these areas, future research can advance EEG-based mental stress detection, making it more reliable and applicable in various domains, ultimately improving mental health and well-being.

ACKNOWLEDGEMENT

We would like to express our deep gratitude to Assoc. Prof Than Quang Khoat, who gives us this golden opportunity to work on this wonderful project. Special thanks to Mr. Nguyen Tuan Dat (IT-E6-04 K67) for the Mindlink EEG headset. Without these tremendous instructions and supports, we cannot complete this project.

8. REFERENCES

[1] N. Sulaiman, M. N. Taib, S. Lias, Z. H. Murat, S. A. M. Aris, M. Mustafa, et al., "Development of EEG-based stress index," in 2012 International Conference on Biomedical Engineering (ICoBE), 2012, pp. 461-466.

[2] Teplan, Michal. (2002). Fundamental of EEG Measurement. MEASUREMENT SCIENCE REVIEW.2.

[3] Perry, A., Saunders, S. N., Stiso, J., Dewar, C., Lubell, J., Meling, T. R., Solbakk, A. K., Endestad, T., & Knight, R. T. (2017). Effects of prefrontal cortex damage on emotion understanding: EEG and

behavioural evidence. Brain, 140(4), 1086–1099. https://doi.org/10.1093/brain/awx031

[4] Sazgar, M., & Young, M. G. (2019). Overview of EEG, Electrode Placement, and Montages. In Springer eBooks (pp. 117–125). https://doi.org/ 10.1007/978-3-030-03511-2_5

[5]. B. Wu, F. Yang, J. Zhang, Y. Wang, X. Zheng and W. Chen, “A frequency-temporal-spatial method for motor-related electroencephalography pattern recognition by comprehensive feature optimization,” Computers in biology and medicine, vol. 42, no. 4, 3530363, April 2012.

[6] Hou, X.; Liu, Y.; Sourina, O.; Tan, Y.R.E.; Wang, L.; Mueller-Wittig, W. EEG based stress monitoring. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 3110–3115.

[7] *Stress*. (2022, June 17). WHO

[8] N. Sulaiman, M. N. Taib, S. Lias, Z. H. Murat, S. A. M. Aris, M. Mustafa, et al., “Development of EEG-based stress index,” in 2012 International Conference on Biomedical Engineering (ICoBE), 2012, pp. 461-466.

[9] S. S. Hassellund, A. Flaa, L. Sandvik, S. E. Kjeldsen, and M. Rostrup, “Long-term stability of cardiovascular and catecholamine responses to stress tests: An 18-year follow-up study,” Hypertension, vol. 55, pp. 131-136, 2010.

[10] S. Ssanghee, G. Yeongjun, and L. Jungtae, “The Relation between Affective Style of Stressor on EEG Asymmetry and Stress Scale during Multimodal Task,” in Third International Conference on Convergence and Hybrid Information Technology, ICCIT '08, pp. 461-466, 2008.

[11] S. H. Seo and J. T. Lee, “Stress and EEG,” Convergence and Hybrid Information Technologies, Marius Crisan (Ed.): InTech, pp. 413-426, 2010.

[12] Coan, J.A. and Allen, J.J.B., “Frontal EEG asymmetry as a moderator and mediator of emotion”, Biol. Psychol., vol. 67, no. 12, pp. 7-49, 2004

[13] Davidson, R.J., 2004. What does the prefrontal cortex “do” in affect: perspectives on frontal EEG asymmetry research. Biological Psychology. 67, 219-233

[14] Davidson, R.J., Jackson, D.C., and Kalin, N.H., “Emotion, plasticity, context and regulation: Perspectives from affective neuroscience”, Psychol. Bull. , vol. 126, no. 6 , pp. 890-906, 2000

[15] Papousek, I. and Schulter, G. 2002. Covariations of EEG asymmetries and emotional states indicate that activity at frontpolar locations is particularly affected by state factors. Psychophysiology. 39(3), 350-60

[16] J. D. Payne, L. Nadel, J. J. B. Allen, K. G. F. Thomas, and W. J. Jacobs, “The effects of experimentally induced stress on false recognition,” Memory, pp. 1-6, 2002.

[17] N. Sulaiman, M. N. Taib, S. Lias, Z. H. Murat, S. A. M. Aris, M. Mustafa, et al., “Development of EEG-based stress index,” in 2012 International Conference on Biomedical Engineering (ICoBE), 2012, pp. 461-466.

[18] Brosschot, J. F., Benschop, R. J., Godaert, G. L., De Smet, M. B., Olff, M., Heijnen, C. J., & Ballieux, R. E. (1992). Effects of experimental psychological stress on distribution and function of peripheral blood cells. Psychosomatic Medicine, 54(4), 394–406. https://doi.org/10.1097/00006842-199207000-00002

[19] Abed, S., & Bagher, M. (2011). Classification of Emotional Stress Using Brain Activity. In InTech eBooks. https://doi.org/10.5772/18294

[20] Aftanas, L.I.; Reva, N.V.; Varlamov, A.A.; Pavlov, S.V. & Makhnev, V.P. (2004). Analysis of Evoked EEG Synchronization and Desynchronization in Conditions of Emotional Activation in Humans: Temporal and Topographic Characteristics. Neuroscience and Behavioral Physiology, Vol. 34, No. 8, pp. 859-867.

[21] Hosseini, S.A. (2009). Quantification of EEG signals for evaluation of emotional stress level, M.Sc. Thesis Report, Biomedical Engineering, Islamic Azad University-Mashhad Branch. (Thesis in Persian).

[22] Kim, K.H.; Bang, S.W. & Kim, S.R. (2004). Emotion recognition system using short-term monitoring of physiological signals, Medical and Biological engineering and Computing, Vol. 42, pp. 419-427.

[23] Alharan, A. F. H., Fatlawi, H. K., & Ali, N. S. (2019). A cluster-based feature selection method for image texture classification. Indonesian Journal of Electrical Engineering and Computer Science, 14(3), 1433. https://doi.org/10.11591/ijeecs.v14.i3.pp1433-1442

[24] Khosrowabadi, R.; Quek, C.; Ang, K.K.; Tung, S.W.; Heijnen, M. A Brain-Computer Interface for classifying EEG correlates of chronic mental stress. In Proceedings of the 2011 International Joint Conference on Neural Networks, San Jose, CA, USA, 31 July–5 August 2011; pp. 757–762.

[25] Sani, M.; Norhazman, H.; Omar, H.; Zaini, N.; Ghani, S. Support vector machine for classification of stress subjects using EEG signals. In Proceedings of the 2014 IEEE Conference on Systems, Process and Control (ICSPC 2014), Kuala Lumpur, Malaysia, 12–14 December 2014; pp. 127–131.

[26] Hou, X.; Liu, Y.; Sourina, O.; Tan, Y.R.E.; Wang, L.; Mueller-Wittig, W. EEG based stress monitoring. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 3110–3115.

[27] Saeed, S.M.U.; Anwar, S.M.; Khalid, H.; Majid, M.; Bagci, A.U. EEG based Classification of Long-term Stress Using Psychological Labeling. *Sensors* 2020, *20*, 1886.

[28] Arsalan, A.; Majid, M.; Butt, A.R.; Anwar, S.M. Classification of perceived mental stress using a commercially available EEG headband. *IEEE J. Biomed. Health Inform.* 2019, *23*, 2257–2264.

[29] Hasan, M.J.; Kim, J.M. A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals. *Brain Sci* 2019, *9*, 376.

[30] Darzi, A.; Azami, H.; Khosrowabadi, R. Brain functional connectivity changes in long-term mental stress. *J. Neurodev. Cogn.* 2019, *1*, 16–41.

[31] Masood, K.; Alghamdi, M.A. Modeling Mental Stress Using a Deep Learning Framework. *IEEE Access* 2019, *7*, 68446–68454.

[32] Camacho, J., & Manian, V. (2016). Real-time single channel EEG motor imagery based Brain Computer Interface. https://doi.org/10.1109/wac.2016.7582973.

[33] Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/bf00116251

[34] Team, 365. (2024, May 15). Introduction to Decision Trees: Why Should You Use Them? 365 Data Science. https://365datascience.com/tutorials/machine-learning-tutorials/decision-trees/