RESEARCH ARTICLE OPEN ACCESS

EEG-Based Stress Level Detection Using Machine Learning

Mrs.C.Sasirekha[1], Kiruthick Raj B J [2], Premkumar J K [3], Kowshik Narayan R [4], Eswaralal B K [5]
1 Assistant Professor, Electronics & Communication Engineering, K.L.N College of Engineering, India
2 UG Scholar, Electronics & Communication Engineering, K.L.N College of Engineering, India
3 UG Scholar, Electronics & Communication Engineering, K.L.N College of Engineering, India
4 UG Scholar, Electronics & Communication Engineering, K.L.N College of Engineering, India
5 UG Scholar, Electronics & Communication Engineering, K.L.N College of Engineering, India

--************************----------------------------------

Abstract

This project introduces a practical and intelligent approach to stress detection using EEG signals and machine learning. By capturing and analyzing brainwave activity, the system identifies five key frequency bands—delta, theta, alpha, beta, and gamma—through the use of a Butterworth bandpass filter. From these bands, meaningful features are extracted to reflect the brain's cognitive and emotional states. These features are then processed through a hybrid pipeline combining unsupervised clustering with supervised learning models, including Support Vector Machine (SVM), Random Forest (RF), and XGBoost, to classify stress levels into Low, Moderate, and High. To bring this solution closer to real-world application, the trained models are integrated into a user-friendly interface built with Streamlit, enabling real-time monitoring and predictions. The result is a robust, automated system capable of supporting mental health assessments in various settings—from clinical environments to personal wellness applications—offering both scalability and accessibility.

Keywords — EEG signal analysis, Machine learning, Stress detection, Mental health monitoring, Real-time deployment, Streamlit.

--************************----------------------------------

1. INTRODUCTION

This project presents an intelligent, automated framework that leverages machine learning to classify stress levels based on EEG brainwave signals. By applying a Butterworth bandpass filter, the system isolates five critical frequency bands—delta, theta, alpha, beta, and gamma—to extract meaningful features that reflect brain activity patterns. The initial classification of stress levels is performed using the K-Means clustering algorithm, which groups data based on signal similarity. These groupings are then refined through supervised classification using algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) to improve prediction accuracy.

To make the system practical for real-time use, the trained models are integrated into an interactive web-based application using Streamlit. This provides a responsive and scalable platform for continuous stress monitoring. The solution has promising applications in areas like workplace stress evaluation, cognitive workload assessment, and tailored mental health support. By automating the process of stress detection, the framework not only supports early diagnosis but also reduces dependency on self-assessment, opening up new possibilities for AI-assisted mental health care.

2. METHODOLOGY

The EEG-Based Stress Detection System is designed to analyze and classify individual stress levels by applying machine learning techniques to EEG data. The process follows a structured workflow that includes data collection, preprocessing, feature extraction, labeling, model training, and finally, deployment through a user-friendly web interface.

2.1 Software Requirements

The system is developed using Python, supported by a suite of powerful libraries and tools commonly used in data science and machine learning. Below is a summary of the software components involved:

- **Python 3.x** – Serves as the primary programming language for data handling, analysis, and application development.
- **NumPy & Pandas** – Essential for manipulating and processing datasets efficiently during each stage of the pipeline.
- **Scikit-learn** – It provides comprehensive support for preprocessing datasets, training various machine learning models, and evaluating their performance through built-in metrics and validation techniques.
- **XGBoost** – Used to build a high-performance gradient boosting classifier optimized for handling structured data.
- **Matplotlib & Seaborn** – Employed to create insightful visualizations such as heatmaps, classification reports, and performance graphs.
- **Joblib** – Joblib is utilized to serialize and deserialize machine learning models, making it easy to store and retrieve them during different stages of deployment or testing.
- **Streamlit** – A lightweight and easy-to-use framework for building an interactive web app that delivers real-time stress predictions.
- **VS Code / Jupyter Notebook** – Used as development environments to write, test, and visualize code and results.
- **CSV Logging** – Helps maintain structured records of processed data, model predictions, and performance metrics.

2.2 Data Acquisition and Preprocessing

For this study, EEG data is sourced from the publicly available SAM-40 dataset, which provides recordings captured under a variety of scenarios designed to evoke cognitive and emotional stress. The dataset offers a reliable foundation for training and evaluating stress detection models.

Preprocessing

Raw EEG signals often contain significant noise and unwanted artifacts that can obscure meaningful patterns. To ensure the clarity and accuracy of the data, a series of preprocessing steps are applied:

Bandpass Filtering: A **Butterworth bandpass filter** is used to clean the EEG signals by removing baseline drift and suppressing external interference. This process focuses on isolating five core EEG frequency bands, each linked to distinct mental states:

- **Delta (0.5–4 Hz)** – Related to deep sleep and unconscious brain activity.
- **Theta (4–8 Hz)** – Often associated with relaxation, meditation, and drowsiness.
- **Alpha (8–13 Hz)** – Reflects a calm, alert state of mind.
- **Beta (13–30 Hz)** – Connected to active thinking, focus, and decision-making.
- **Gamma (30–100 Hz)** – Typically associated with higher-order cognitive activities such as learning, attention, and memory recall.

The filtered signals obtained from this step form the foundation for extracting meaningful features that help in stress level classification.

2.3 Feature Extraction

After the EEG signals are filtered, the next step involves extracting meaningful features that can be used to assess mental stress levels. For this, **Power Spectral Density (PSD)** features are calculated from each EEG segment. the average power within each of the five key frequency bands Delta, Theta, Alpha, Beta, and Gamma is measured.

These power values act as numerical markers that reflect the subject's cognitive state and mental workload. As a result, every EEG data sample is converted into a feature vector containing five distinct values, each representing the power level in one of the respective frequency bands. This structured representation is then used for stress classification in the next stage of the pipeline.

2.4 Model Development

The process of building the predictive model is carried out in two key stages: assigning stress levels through unsupervised learning and then training classifiers to recognize those levels using supervised techniques.

2.4.1 Data Labeling via Clustering

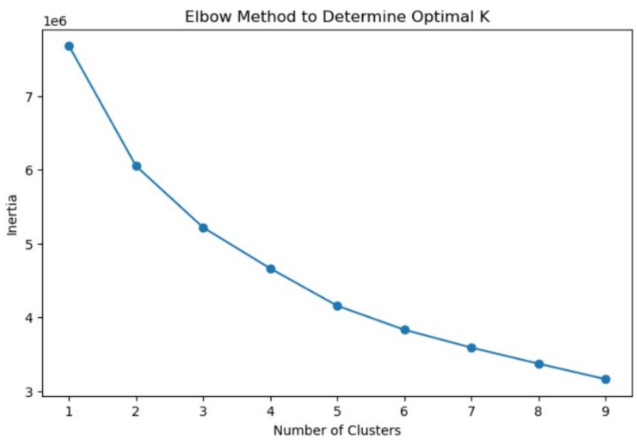

Fig 1. Elbow method to determine optimal k

Since the dataset does not include predefined labels for stress intensity, an unsupervised learning method is used to identify patterns in the data. Specifically, **K-Means clustering** is applied to the feature vectors derived from EEG signals to group the samples into three distinct stress categories:

- **Low Stress**
- **Moderate Stress**
- **High Stress**

To determine the optimal number of clusters (k), the **Elbow Method** is employed, which helps visualize where adding more clusters no longer significantly improves clustering performance (as illustrated in *Figure 1*).

The cluster centers (centroids) are then interpreted based on the frequency band distribution. Patterns dominated by higher power in the **Beta** and **Gamma** bands are associated with higher stress levels, while those with greater power in the **Delta** and **Theta** ranges are linked to relaxed or low-stress states. This clustering strategy provides a biologically meaningful way to label the data for further classification.

2.4.2 Machine Learning Classification

Once the dataset is labeled through clustering, it is used to train and evaluate a set of supervised machine learning models aimed at accurately classifying stress levels. The following algorithms are employed:

- **Support Vector Machine (SVM)** – Uses an RBF kernel to handle non-linear data.
- **Random Forest (RF)** – Combines multiple decision trees to improve accuracy and reduce overfitting.
- **XGBoost** – A fast and efficient gradient boosting algorithm suited for structured data.

The labeled data is split into training and testing sets using an 80:20 ratio. Model performance is assessed using a comprehensive set of evaluation metrics, including **accuracy**, **precision**, **recall**, **F1-score**, and the **confusion matrix**. These metrics help measure not just overall performance but also how well the models distinguish between the different stress categories.

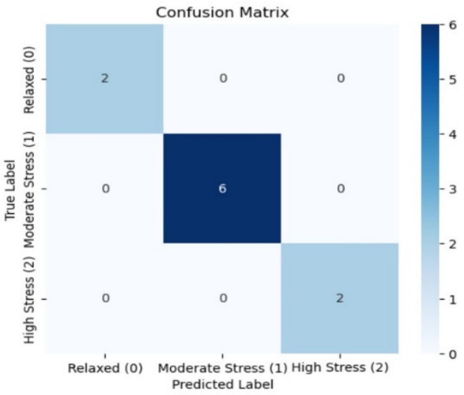

Fig 2. Confusion Matrix

2.4.3 Feature Standardization

Before training, all input features are scaled using Scikit-learn's **StandardScaler** to ensure they have a uniform distribution. This normalization step helps prevent features with larger values from dominating the model and reduces bias due to variance differences, ultimately improving classification performance.

2.5 Deployment: Web-Based Stress Prediction Interface

A lightweight and interactive web application is built using **Streamlit** to provide real-time stress predictions. Users can input EEG power values **Delta, Theta, Alpha, Beta, and Gamma** either manually or by uploading a file.

Once submitted, the trained model instantly analyzes the input and displays the predicted stress level on-screen. Designed to be simple and responsive, the interface is accessible to both technical and non-technical users, making it ideal for use in wellness centers, research settings, or for personal stress monitoring.

2.6 Flow Diagram

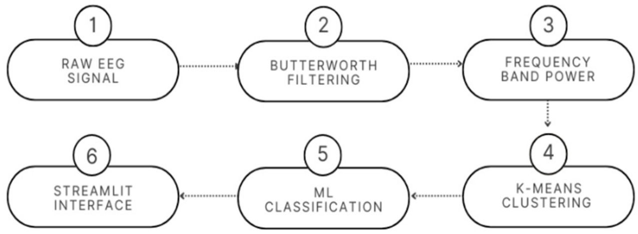

Fig 3. Flow Diagram

3. Result

The EEG-Based Stress Detection System turned out to be both reliable and practical when it came to recognizing stress levels based on brainwave activity. Using data from the publicly available SAM-40 dataset—recorded while participants were exposed to various mentally and emotionally demanding tasks—the system begins by cleaning and filtering the raw EEG signals. It isolates five core brainwave frequencies: Delta, Theta, Alpha, Beta, and Gamma. Each of these bands tells us something different about the person's mental state, from deep relaxation and calmness to intense focus or high cognitive load. To make sense of this information, the system calculates the average power within each frequency band using Fast Fourier Transform (FFT), creating a simplified and structured profile of brain activity for each individual session.Since the dataset didn't include direct stress level labels, the system first used K-Means clustering—an unsupervised learning approach—to group similar brainwave patterns into three natural categories: low, moderate, and high stress. These clusters acted as the foundation for the next stage, where supervised machine learning models were trained to recognize and predict these stress levels. The models included Support Vector Machine (SVM), Random Forest (RF), and XGBoost—each tested using an 80:20 training-to-testing split. The performance of each model was carefully evaluated using metrics like accuracy, precision, recall, and F1-score. Among them, the SVM model came out on top, delivering the highest accuracy and showing consistent results across the board. That said, both Random Forest and XGBoost also performed well, proving that the extracted EEG features were informative and dependable for stress detection.

What makes this system even more impactful is how it brings advanced AI into a simple, real-world experience. A web-based application was built using Streamlit to allow users to interact with the model in real time. Rather than needing to upload large EEG files or have technical knowledge, users can simply enter five values—each representing the average power of a frequency band—and get an instant prediction of their stress level. The interface is clean, responsive, and designed to be intuitive for anyone to use, whether in a clinical setting, a workplace, or even at home for personal mental health tracking.

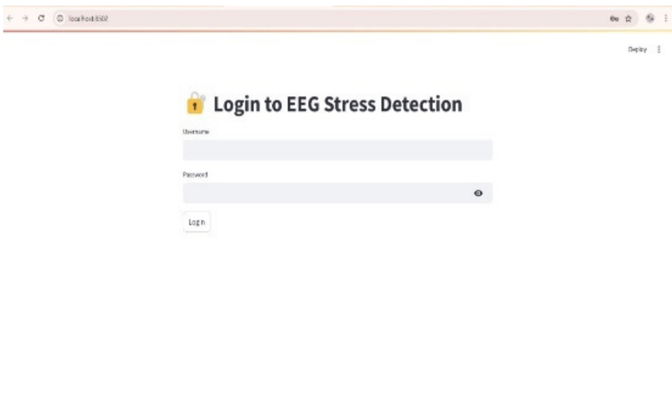

Fig 4. Website Login Page

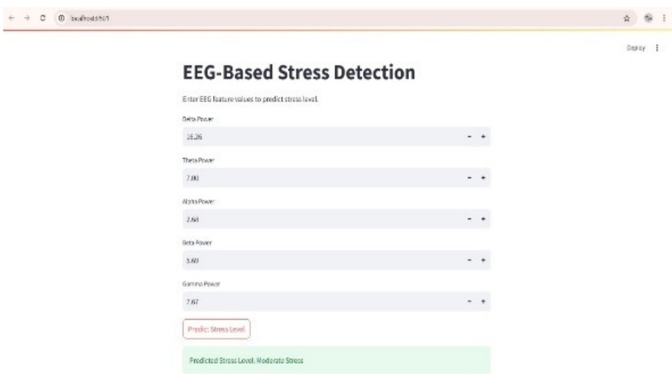

Fig 5. Website Result Interface

This all-in-one approach—combining signal processing, unsupervised learning, supervised classification, and real-time deployment—shows just how practical and powerful EEG-based stress detection can be. It opens up new possibilities for non-invasive mental health monitoring, offering early insights into stress before it becomes overwhelming. Whether used by professionals or individuals, this system provides a scalable and accessible tool to help people better understand and manage their mental well-being in everyday life.

4. CONCLUSIONS

In conclusion, the EEG-based stress detection system developed in this study presents a robust and intelligent approach to monitoring mental health in real time. By extracting meaningful power features from core EEG frequency bands—Delta, Theta, Alpha, Beta, and Gamma—and applying machine learning techniques, the system effectively classifies stress into three distinct levels: Low, Moderate, and High. The use of K-Means clustering for initial data labeling, followed by classification using models like Support Vector Machine (SVM), Random Forest, and XGBoost, demonstrates strong and consistent performance across varying conditions and individuals. What sets this system apart is not only its accuracy but also its accessibility. The lightweight, user-friendly web interface built with Streamlit allows users to interact with the model in a simple, intuitive way—making stress assessment possible even outside of clinical or research environments.

This work also opens the door to exciting future possibilities. With further improvements such as the integration of deep learning algorithms, larger and more diverse EEG datasets, and the use of real-time wearable devices, the system could evolve into a continuous stress tracking solution. Such enhancements would make it even more responsive and adaptive, enabling early intervention and personalized mental wellness strategies. Ultimately, this project highlights the real potential of EEG and AI integration in supporting mental health—offering a scalable, non-invasive, and practical tool for proactive stress management and emotional well-being.

6. REFERENCES

[1] Vanitha V, Krishnan P "Real time stress detection system based on EEG signals" Biomedical Research 2016; Special Issue: S271-S275.

[2] Ankita Gandhi J ,Dr. Udesang K. Jaliya Stress "Detection through EEG Signals: Employing a Hybrid Approach integrating Time Domain, Frequency Domain Features and Machine Learning Techniques" Electrical Systems 20-3 (2024):3965 – 3973.

[3] TuerxunWaili1, Yousif, Sa'ad Alshebly2, Khairul Azami Sidek2 and Md Gapar Md Johar1 "Stress recognition using Electroencephalogram (EEG) signal" ICTEC 2019 Journal of Physics: Conference Series.