Project Report: EEG-Based Attention Classification Using Machine Learning and Deep Learning

Introduction

This project focuses on classifying attention states using EEG (Electroencephalography) data obtained from the EMOTIV device. The goal is to build a machine learning pipeline that processes EEG signals to classify different mental states, such as focused, unfocused, and drowsy, using traditional machine learning techniques such as SVM (Support Vector Machine) and more advanced techniques like decision trees and random forests. The project also incorporates a deep learning-based approach to optimize the classification.

Dataset Overview

The dataset consists of EEG data collected across 34 experiments. Each experiment includes data from several channels corresponding to different brain regions. The dataset provides a high sampling rate (128 Hz) and captures different brain wave bands (alpha, beta, gamma, delta).

- **Channels:** AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4.
- **Bands:**
    - Alpha (8-13 Hz)
    - Delta (0.5-4 Hz)
    - Beta (13-30 Hz)
    - Gamma (30+ Hz)

Pipeline

1. **Data Loading and Preprocessing**:
    - EEG signals were loaded from `.mat` files using `scipy.io` functions. The signals were segmented into different states based on the timestamp: focused, unfocused, and drowsy.
    - A preprocessing pipeline was built using `scikit-learn` for standardization of the EEG signals.
2. **Modeling Techniques**: Several machine learning models were trained on the EEG dataset:

SVM (Support Vector Machine): Initially, SVM was applied on scaled EEG data from individual participants. The dataset was split into training and test sets with an 80-20 split.

- **Results**: The SVM model achieved good accuracy in distinguishing between different attention states when tested on both individual and multiple participant datasets.

Decision Trees and Random Forests: These models were applied for better generalization across the entire dataset.

○

Random Forest Results: The depth of trees in the Random Forest model was optimized using cross-validation, with accuracy plotted for varying depths.

■

K-Fold Cross-Validation: A 5-fold cross-validation strategy was used to avoid overfitting and improve model robustness.
python
Copy code

○

3. **Deep Learning Approach**: A TensorFlow-based model was developed as an alternative approach for more complex pattern recognition in the EEG signals. The model was designed to handle multivariate time-series data, taking advantage of the periodic nature of EEG signals.
    - **Results**: The deep learning model showed improvement over traditional methods but required more computational resources and careful tuning to avoid overfitting.

Integration with GUI

The trained models were saved and then integrated into a Tkinter-based GUI. The GUI allows users to upload `.mat` files, visualize EEG signals, and get real-time predictions of attention states.

- **Key GUI Features**:
    - Upload and plot EEG data from `.mat` files.
    - Display power levels for different brain wave bands.
    - Show classification results based on trained models.

Challenges

- Handling large datasets with multiple channels and segments was challenging, and required efficient preprocessing and memory management.
- The EEG data is noisy, making it difficult to achieve high accuracy without extensive filtering and feature extraction.

Future Work

- Extend the deep learning approach using Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models to better capture temporal dependencies in EEG signals.
- Experiment with feature engineering to improve model performance by extracting more meaningful features from the EEG signals.

Conclusion

This project successfully demonstrates the application of machine learning and deep learning to classify attention states using EEG signals. The GUI integration provides a user-friendly way to visualize EEG data and make predictions using pre-trained models. Future improvements could further enhance model accuracy and expand the scope of the application to real-time EEG monitoring.