Assignment: EEG-Based Emotion Recognition Using Neural Networks

Introduction

In this assignment, you will explore **EEG-based emotion recognition** by using neural networks to classify emotions and stress levels from EEG data. The dataset used for this project is the **SAM 40 dataset**, which consists of EEG recordings from 40 participants (average age: 21.5) during various cognitive tasks (Stroop test, arithmetic, symmetry recognition, and relaxation). This dataset is helpful for analyzing cognitive stress levels in response to different mental tasks. The goal of this assignment is to develop a model that can classify emotional states using this EEG data.

Resources

1. **Scientific Article:**
    - EEG-Based Emotion Recognition Using Neural Networks (ScienceDirect)
2. **Dataset:**
    - SAM 40 Dataset (Figshare)
3. **EEG Basics Resource:**
    - Introduction to EEG (AlexEnge)

Key Tasks to Complete

1. Download and Explore the Dataset

- Download the SAM 40 dataset from Figshare.
- Inspect the raw and processed EEG data.
- Understand the structure of the dataset: raw vs. filtered data, task segmentation, and artifact removal.

2. Perform Exploratory Data Analysis (EDA)

- Conduct an **exploratory data analysis (EDA)** to understand the distribution and characteristics of the EEG data.
- Create plots such as:
    - **Time-domain plots**: Plot the EEG signals over time to identify patterns.
    - **Frequency-domain analysis**: Perform Fourier transforms and visualize power spectral densities (PSDs) to explore the frequency content of EEG data.
    - **Topographic maps**: Visualize the spatial distribution of EEG signals across the 32 channels.

    - **Task-specific EEG plots**: Demonstrate how EEG signals vary across different cognitive tasks (e.g., Stroop test, arithmetic task).

3. Perform Basic EEG Data Analysis Techniques

- Apply basic **EEG analysis techniques** such as:
    - **Band Power Calculation**: Calculate the power of different frequency bands (e.g., delta, theta, alpha, beta) to explore how they vary with cognitive stress.
    - **Independent Component Analysis (ICA)**: Use ICA to remove artifacts from EEG data, such as eye blinks or muscle artifacts.
    - **Spectral Analysis**: Perform spectral decomposition (e.g., via Fast Fourier Transform or wavelet transform) to identify frequency bands associated with specific cognitive states.

4. Setup the Development Environment

- Install necessary Python dependencies, including PyTorch, for neural network implementation.

5. Data Loading and Preparation

- Implement the `load_dataset`, `load_labels`, and `split_data` functions to properly load and process EEG data.
- Use the `EEGDataset` class for efficient batch processing during training.
- Preprocess the EEG signals for model training.

6. Model Development

- **Model 1: SimpleNN** - A simple multi-layer perceptron for classification.
    - Train the model to recognize emotions and stress levels from the EEG data.
- **Model 2: EEG_CNN** - A Convolutional Neural Network for more advanced feature extraction from EEG signals.
    - Compare the performance of both models.

7. Model Training

- Use functions like `train_one_epoch` and `validate` for training and evaluation.
- Implement early stopping using `EarlyStopping` to prevent overfitting.
- Save the best-performing model for later use.

8. Evaluation and Insights

- Assess the performance of both models on the validation set.
- Identify any challenges faced during training, such as noise in the data or subject variability.
- Suggest possible improvements for future iterations.

9. Model Optimization (Optional)

- Tweak hyperparameters and model architecture to improve accuracy.
- Implement more advanced preprocessing techniques for better artifact removal.

10. Documentation and Submission

- Provide clear and well-documented code and analysis.
- Submit the following as part of your assignment:
    - A **well-documented GitHub repository** containing all the code with detailed comments and explanations.
    - A **Notion page** that includes the analysis, experiments, results, and findings. Ensure that the analysis is clearly laid out, and all experiments are documented in detail.
    - A **presentation** summarizing the work and results, to be presented to the team. The presentation should be clear, concise, and aimed at explaining the approach, model performance, challenges, and key insights.
- Ensure that all code is thoroughly commented and explained. Do not use online code for similar tasks; if you reference code, make sure to understand it fully and explain each line.

Rules for the Assignment:

1. **Understand and Explain the Code**:
    - If you refer to or use code snippets from online resources, ensure you fully understand the logic behind the code. Explain each line of the code in your documentation and your repository.
    - **Do not copy and paste code without understanding**.
2. **Well-Documented Code**:
    - Write clean, well-documented, and easily understandable code. Comment every function and section explaining its purpose and functionality.
3. **In-Depth Analysis**:
    - Your analysis should be thorough, including any issues you encountered, possible improvements, and performance metrics. This should be included in your **Notion page**.
4. **Presentation**:
    - Create a professional presentation summarizing the steps, results, and insights from your work. This should be in a format that could be presented to a technical team.
5. **Timeline**:
    - You are given **one week** to complete this assignment. Please ensure that you manage your time effectively and submit the work within the provided timeframe.

Dataset Description

The **SAM 40 dataset** consists of EEG time-series data recorded during various cognitive tasks. Each subject's EEG was recorded using a **32-channel Emotiv Epoc Flex** gel kit with a sampling rate of **128 SPS (1024 Hz internal)**. The dataset is divided into raw and filtered versions, where raw data contains noise and artifacts, and filtered data is cleaned and ready for analysis.

Key Features of the Dataset:

- **Number of Subjects**: 40 participants
- **Number of Channels**: 32 EEG channels
- **Tasks**: Stroop test, arithmetic task, mirror image recognition, and relaxation
- **Feedback**: The dataset includes a file containing the stress ratings from participants (scales.xls).
- **Artifact Removal**: MATLAB code is provided to help with artifact removal.

You can access the dataset and download the raw or filtered EEG data using the following link to Figshare.

Expected Deliverables

1. **Code**:
    - The code for loading, preprocessing, and training the models should be well-commented and organized.
2. **Report**:
    - Provide a report that includes:
        - A brief description of the dataset.
        - A summary of your model development process.
        - The results of your experiments and an analysis of the challenges faced during training.
        - Possible improvements for future work.
3. **Visualizations**:
    - Include any plots or graphs that help in understanding the model's performance, such as loss curves, accuracy plots, or confusion matrices. Additionally, include the visualizations from your **Exploratory Data Analysis (EDA)** phase (e.g., time-domain plots, frequency-domain analysis, and topographic maps).
4. **Performance Metrics**:
    - Provide detailed performance metrics (accuracy, precision, recall, etc.) of the trained models.

Links to Resources and Dataset

- **Scientific Article on Emotion Recognition**: ScienceDirect
- **SAM 40 Dataset (Figshare)**: Download SAM 40 Dataset
- **EEG Basics Resource**: Introduction to EEG (AlexEnge)