Project Overview

- **Project Title:** Time Series Analysis
- **Main Objective:** Stress and Affect Detection
- **Author:** Rian Renold Dbritto
- **NUID:** 002026598

Data Used

The project utilizes data from the WESAD (Wearable Stress and Affect Detection) dataset. Specifically, it loads `.pkl` files for multiple subjects (S2, S4, S6, S8, S10) from a Google Drive path (`/content/drive/MyDrive/WESAD FILES/`).

Each `.pkl` file appears to contain a dictionary with the following main keys:

- `signal`: Contains sensor data from both `chest` and `wrist` devices.
    - **Chest Data**: Includes ACC (Accelerometer), ECG (Electrocardiogram), EMG (Electromyography), EDA (Electrodermal Activity), Temp (Temperature), and Resp (Respiration) signals.
    - **Wrist Data**: Includes ACC, BVP (Blood Volume Pulse), EDA, and TEMP signals.
- `label`: Contains an array of integer labels.
- `subject`: Identifies the subject (e.g., 'S2', 'S4').

The project processes these pickled data files, extracts the chest signals, and verifies their lengths and shapes.

Methodology

The project follows these key steps:

1. **Importing Libraries**: Essential libraries like `pickle`, `numpy`, and `pandas` are imported for data handling and numerical operations.
2. **Data Loading**: Five `.pkl` files (S2, S4, S6, S8, S10) are loaded into `data1` through `data5` variables using `pickle.load()`.
3. **Data Verification**: The loaded data's structure and the lengths/shapes of ECG signals and labels for each subject are checked.
    - For instance, ECG signal for S2 has a length of 4,255,300, and its labels have a shape of (4,255,300,).
4. **Data Transformation to Tabular Form**:
    - Accelerometer (ACC) data from the chest (`acc_x`, `acc_y`, `acc_z`) for each subject is converted into pandas DataFrames (`dfS1` to `dfS5`).
    - A function `add_multiple_columns` is defined to append other chest signals (ECG, EMG, EDA, Temp, Resp) to these DataFrames.

    - Labels and subject IDs are then added as new columns (label and subject) to their respective DataFrames.

Current Status

At the point of analysis, the notebook has successfully loaded the raw WESAD data, extracted relevant chest and label data, and transformed it into a structured tabular format (Pandas DataFrames) for further analysis. For example, dfS2 (data for subject S4) is shown to have 4,496,100 rows and 10 columns, including acc_x, acc_y, acc_z, ECG, EMG, EDA, Temp, Resp, label, and subject.
The notebook sets up the foundation for time series analysis aimed at stress and affect detection by preparing the multi-modal sensor data.

Conclusion

In conclusion, the project successfully initiated the process of Stress and Affect Detection by:

- Loading and verifying raw multi-modal sensor data from the WESAD dataset.
- Extracting and structuring relevant chest signals (ACC, ECG, EMG, EDA, Temp, Resp) and associated labels.
- Transforming this data into a ready-to-use tabular format (Pandas DataFrames), establishing a solid foundation for subsequent time series analysis, feature engineering, and model development.

The project is currently in the data preparation phase, with the data now organized for further analytical steps to achieve the goal of stress and affect detection.