Task #1 Human Activity Recognition

1. Deep Learning Models

Preprocessing

- Used **raw accelerometer and gyroscope data** from the UCI HAR dataset.
- Reshaped the data into **(num_samples, window_size, num_channels)** format for deep learning models.

LSTM

Architecture:

- Two **LSTM layers** (64 units each) with **Dropout (0.3)** to prevent overfitting.
- **Dense layer** with Softmax activation for classification into 6 activity classes.

Training Details:

- **Optimizer:** Adam
- **Loss Function:** Sparse Categorical Crossentropy
- **Epochs:** 15, **Batch Size:** 32
- **Validation Split:** 20% of the training data.

Performance:

- **Training Accuracy:** 84.97%
- **Test Accuracy: 90.13%**

CNN

- **Architecture:**
  - Two **Conv1D layers** (64 and 128 filters) with **ReLU activation**.
  - **MaxPooling1D** layers for dimensionality reduction.
  - **Flatten Layer** followed by **Dense layers**.
  - **Dropout (0.3)** to prevent overfitting.
- **Training Details:**
  - **Optimizer:** Adam

    - **Loss Function:** Sparse Categorical Crossentropy
    - **Epochs:** 15, **Batch Size:** 32
    - **Validation Split:** 20%
- **Performance:**
    - **Training Accuracy:** 85.68%
    - **Test Accuracy: 90.33%**
    - **Observation:** CNN learns local spatial patterns in sensor signals, achieving similar accuracy to LSTM but training faster.

OBSERVATION

CNN and LSTM gave almost same accuracy about 90%

2. ML models

Feature Extraction using TSFEL:

- Extracted features from time-series windows using **`tsfel.time_series_features_extractor()`.**
- Used **`tqdm`** to track progress as feature extraction was taking a long time.
- Set sampling frequency = 50Hz to match the dataset.
- Stored extracted features as **`X_train_features`** and **`X_test_features`.**

Training Machine Learning Models:

- Trained Random Forest, SVM, and Logistic Regression models on the extracted TSFEL features.
- Measured accuracy on test data to compare model performance.

Comparison with Author's Features:

- Loaded pre-extracted features from the UCI HAR dataset (provided by the authors).

- Trained Random Forest using these features and compared the accuracy with TSFEL-extracted features.

Observations

1. Random Forest Accuracy - 92.80%
2. SVM accuracy was - 81.33%
3. Logistic regression - 88.22%
4. Cnn - 90.33 %
5. Lstm - 90.13 %
6. Using authors features and using random forest - 92.46%

- Extracting features took very long
- Random Forest with TSFEL-extracted features was the best-performing model.
- Deep Learning models were effective but didn't outperform ML models significantly.
- TSFEL-extracted features worked slightly better than the dataset authors' precomputed features, but some deviation of accuracy is expected