SUMMARY REPORT

Dataset used: UCI - HAR

Methods tried:

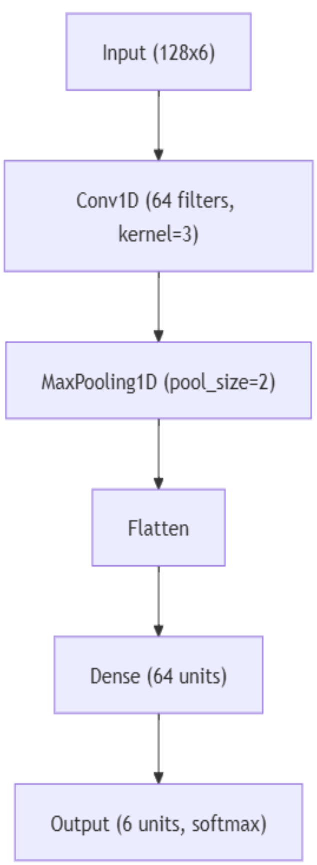
I used the raw inertial signals data such as accelerometer and gyrometer data values, from which we train three deep learning algorithms. Further, TSFEL (Time Series Feature Extraction Library) was used to extract all domain features from the dataset. Those features were bulk processed so as to optimize learning. All features, original, TSFEL and extracted, were scaled for smooth processing.

Architectures:

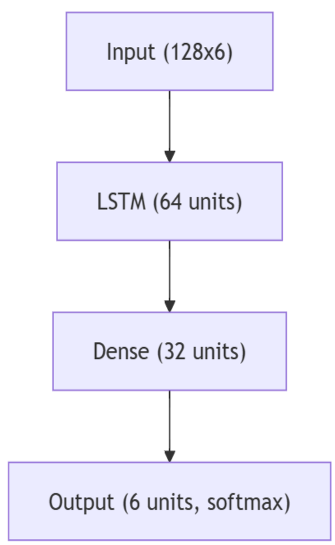
1. DEEP LEARNING : *CNN, LSTM, CNN-LSTM* on features taken from raw inertial signals data.
2. MACHINE LEARNING: *Random forest, SVM and Logistic Regression* on both original 561 features from the UCI HAR dataset and on extracted TSFEL features.

The architectures of deep learning models are as shown:

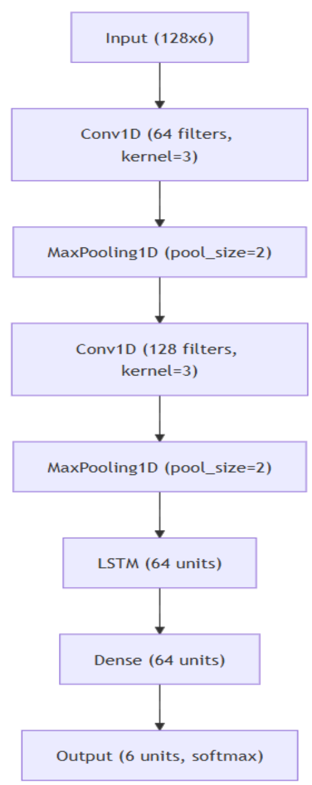
CNN:



LSTM:

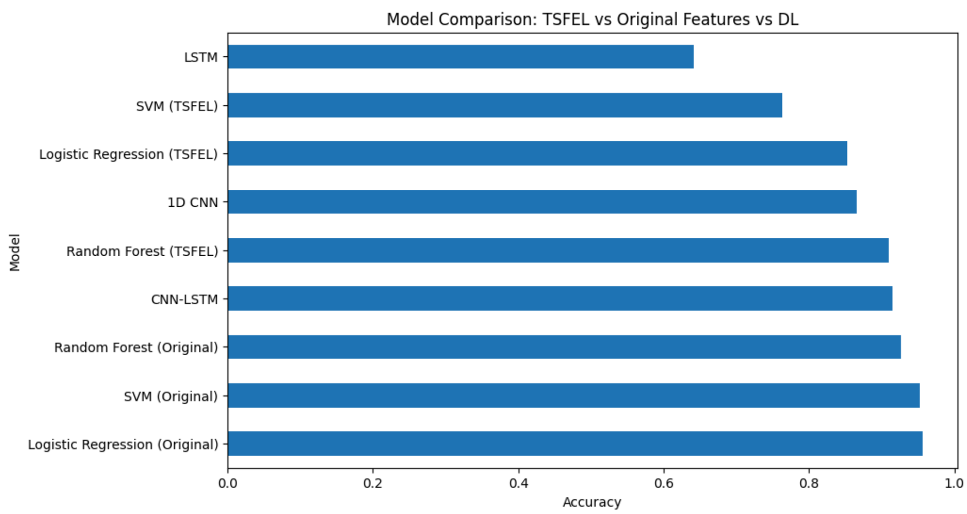


CNN-LSTM:



Results:

Model	Accuracy (Original Features)	Accuracy (TSFEL Features)
Logistic Regression	95.66%	85.31%
SVM	95.22%	76.35%
Random Forest	92.64%	91.01%
CNN-LSTM	91.48%	-
1D CNN	86.53%	-
LSTM	64.10%	-



Observations:

1. Random forest is a robust classifier as it performs well on both original and TSFEL features.
2. Overall, original features for ML models perform the best when compared with their performances over TSFEL features. This may be because The original 561 features are carefully engineered and include domain-specific transformations (e.g., gravity separation, jerk signals). Whereas the TSFEL features ack domain-specific transformations and may miss critical patterns in the data. While Random Forest performs well on TSFEL features (91.01%), SVM and Logistic Regression struggle.
3. DL models (CNN and CNN-LSTM) generalise well even from extracted features from raw signals, upto 90 percent.