Subject-wise Human Activity Recognition using CNN-BiLSTM on Wearable Sensor Data

Islam K M Mozaddedul
228801154

Abstract

Human Activity Recognition (HAR) is a fundamental problem in pervasive computing and healthcare applications, aiming to classify human activities from wearable sensor data. In this project, a multi-sensor time-series dataset containing accelerometer and physiological signals collected from multiple subjects was analysed. A complete machine learning pipeline was implemented, including data preprocessing, subject-wise train-test splitting, time-series segmentation, and deep learning model development.

A CNN-BiLSTM architecture was employed to capture both spatial sensor dependencies and temporal activity patterns. The model was evaluated using accuracy, precision, recall, and F1-score under a strict subject-independent evaluation protocol. Comprehensive visualizations were used throughout the analysis to support preprocessing decisions and interpret model behaviour. The final model achieved an overall test accuracy of 78.4%, demonstrating robust generalization across unseen subjects while highlighting the challenges of distinguishing motion-similar activities.

1. Introduction

Human Activity Recognition (HAR) focuses on identifying physical activities performed by individuals using sensor data collected from wearable devices. Accurate HAR systems are critical for applications such as health monitoring, rehabilitation, smart environments, and sports analytics.

This project utilizes a multi-subject wearable sensor dataset containing inertial measurement unit (IMU) signals recorded from hand, chest, and ankle locations. The dataset includes multiple daily activities with varying motion characteristics.

The objective of this work is to develop a subject-independent HAR model using deep learning techniques. Model performance is evaluated using accuracy, precision, recall, and F1-score. This report is organized as follows: Section 2 presents exploratory data analysis, Section 3 describes data reparation, Section 4 details model training, Section 5 provides mathematical formulation, Section 6 discusses results, and Section 7 concludes the report.

2. Exploratory Data Analysis (EDA)

This section explores the characteristics of the dataset using visualizations to understand activity distribution, sensor behavior, and inter-sensor relationships.

Figure 1: Activity Distribution

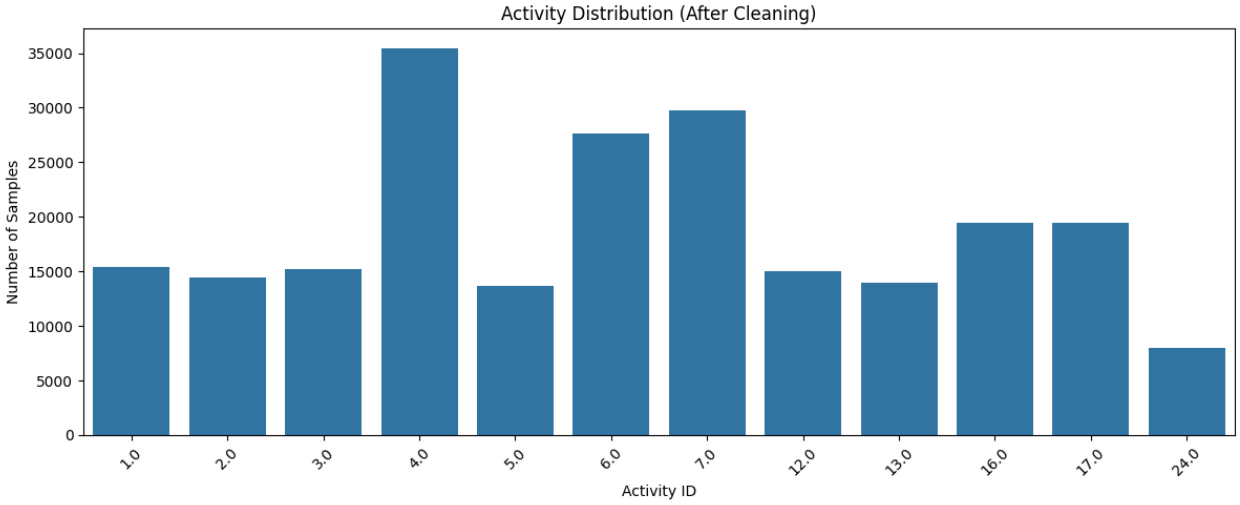


Figure 1: Activity Distribution After Cleaning.

Figure 1 illustrates the distribution of activity samples across different activity classes after preprocessing. A clear class imbalance is observed, where dynamic activities such as walking and cycling contain more samples compared to stair-related and household activities. This imbalance indicates that classification performance may vary across activities and motivates the use of advanced temporal models rather than simple classifiers.

Figure 2: PCA Projection of Sensor Signals

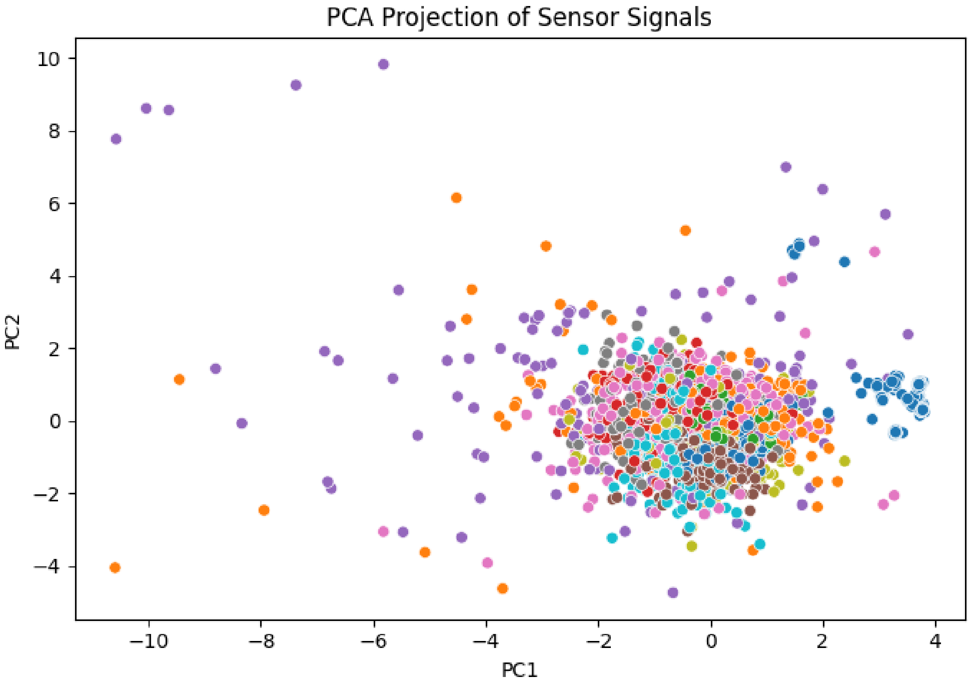


Figure 2: PCA Projection of Sensor Signals.

Figure 2 visualizes the high-dimensional sensor signals projected onto two principal components using PCA. Considerable overlap among activity classes is evident, indicating that several activities share similar motion patterns in the raw feature space. This overlap explains the difficulty of achieving perfect classification and motivates the use of temporal deep learning models rather than simple classifiers.

Figure 3: Raw Sensor Signal Visualization

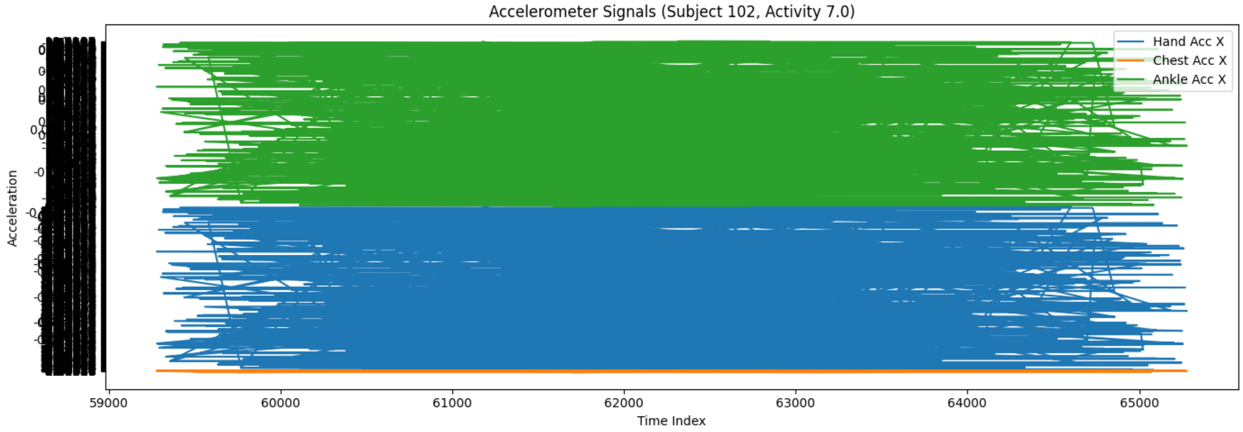


Figure 3:Raw Accelerometer Signal Visualization from Multiple Sensors.

Figure 3 presents raw accelerometer signals collected from hand, chest, and ankle sensors during a walking activity. Periodic patterns can be clearly observed, especially in the ankle sensor, which captures repetitive leg movement. Differences in signal amplitude and frequency across sensor locations highlight the importance of multi-sensor fusion for effective activity recognition.

Figure 4: Sensor Correlation Heatmap

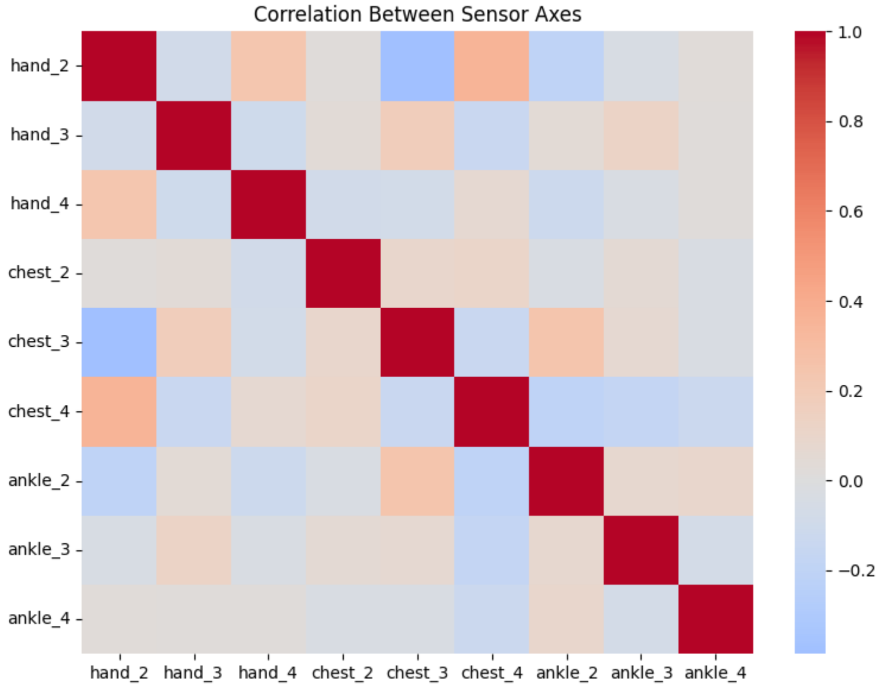


Figure 4: Sensor Correlation Heatmap.

Figure 4 shows the correlation among selected accelerometer axes from different sensor locations. Moderate correlations are observed, indicating that while some redundancy exists, each sensor provides complementary information. This observation supports the use of convolutional layers to automatically learn spatial relationships among sensor channels.

Figure 5: Distribution of Hand Acceleration (X-axis) Across Activities

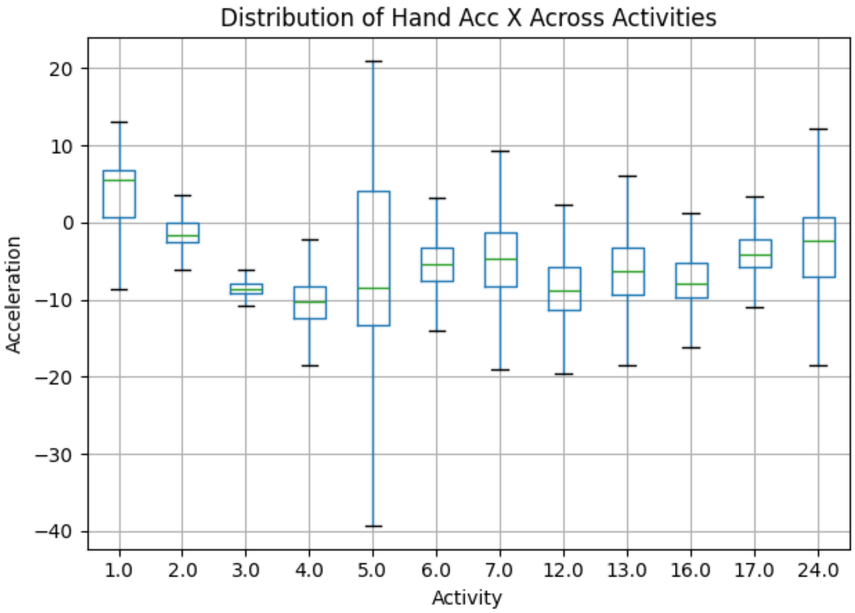


Figure 5: Distribution of Hand Acceleration (X-axis) Across Activities.

Figure 5 illustrates the distribution of hand accelerometer signals along the X-axis for different activities. Distinct medians and variability ranges indicate that activities exhibit unique motion characteristics. However, overlapping distributions suggest challenges in discriminating motion-similar activities such as standing and sitting.

Figure 6: Mean Hand Acceleration Magnitude per Activity

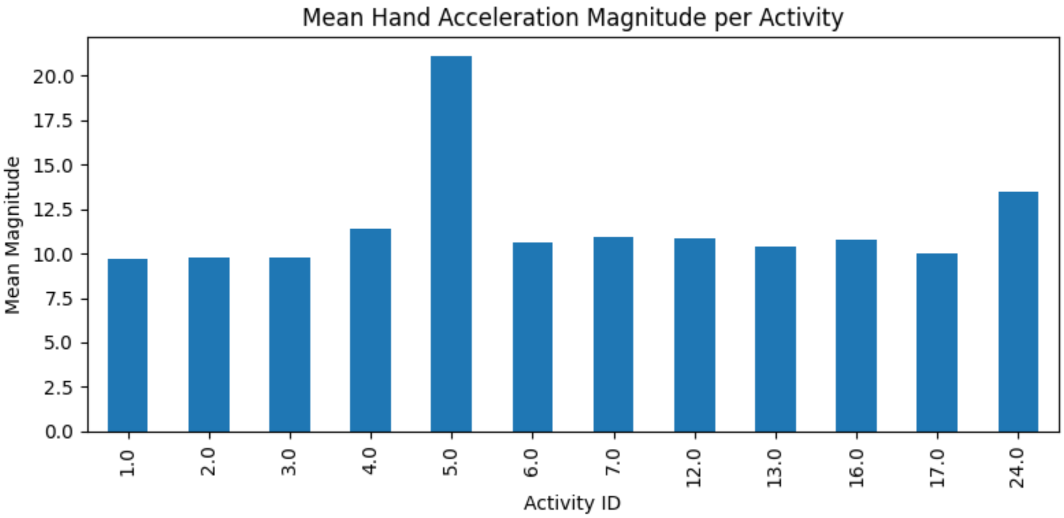


Figure 6: Mean Hand Acceleration Magnitude per Activity.

Figure 6 shows the mean hand acceleration magnitude for each activity. Dynamic activities such as running and rope jumping exhibit higher magnitudes, while static activities such as sitting and standing show lower values. This trend helps explain why dynamic activities are generally classified with higher accuracy.

3. Data Preparation

Data preprocessing plays a critical role in achieving reliable model performance.

Missing values caused by wireless transmission loss and mismatched sampling rates were handled using forward and backward filling within each subject, preserving temporal continuity while preventing data leakage. A strict subject-wise train-test split was applied, ensuring that no subject appeared in both training and test sets.

Time-series segmentation was performed using a sliding window approach with a window length of 200 time steps. This window size was selected to balance temporal context and dataset size.

Figure 7: Window Size vs Sample Count Trade-off

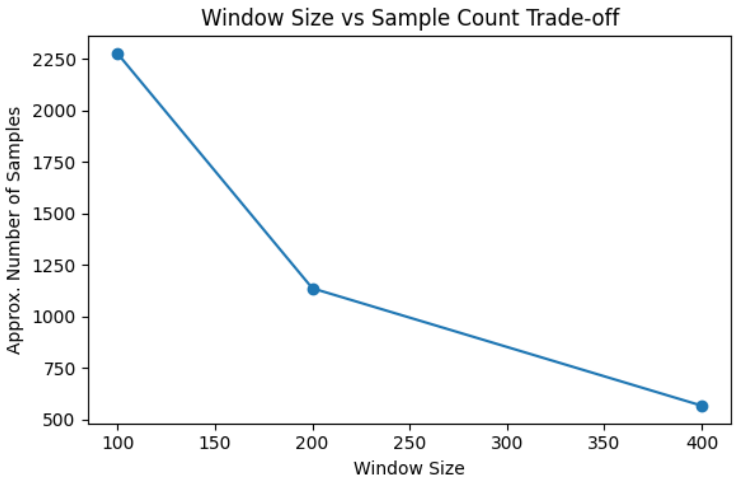


Figure 7: Window Size vs Sample Count Trade-off.

Figure 7 illustrates the trade-off between window size and the number of generated samples. Larger windows provide richer temporal information but significantly reduce the number of training samples. Based on this trade-off, a window size of 200 was selected.

4. Training

A deep learning approach was adopted to model the complex temporal and spatial dependencies in the sensor data.

The CNN-BiLSTM architecture was selected after exploratory experimentation due to its ability to combine spatial feature extraction and temporal sequence modeling. Convolutional layers were used to learn local sensor patterns, while the Bidirectional LSTM layers captured long-term temporal dependencies in both forward and backward directions.

The model was trained using the Adam optimizer with a learning rate of 0.001 and cross-entropy loss. Training was performed for 10 epochs on a GPU-enabled environment.

Figure 8: Training Loss Curve

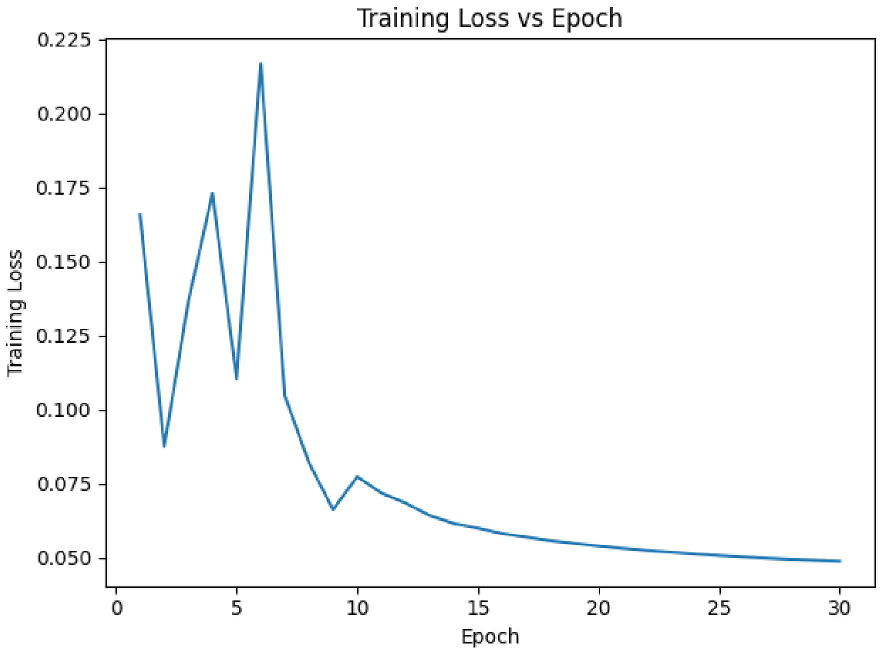


Figure 8: Training Loss Curve of the CNN–BiLSTM Model.

Figure 8 shows the training loss across epochs. The steadily decreasing loss indicates stable convergence and effective learning without severe overfitting.

Table 1: Model Configuration

Component	**Configuration**
Convolution Layers	Conv1D (64, 128), ReLU, MaxPooling
LSTM	Bidirectional LSTM, 2 layers, hidden size 128
Optimizer	Adam
Loss Function	Cross-Entropy
Epochs	30

For tables, place a title with number and use references to the table using hyperlinks as Table 1. Follow the same stylings used in Table 1 – bold the column headers (also the row headers if necessary) and don't use any borders. Please have at least one table to provide detailed results of either the training or the results section. You can choose to use multiple tables in both sections. **Tables are the third most important component in this report**.

5. Mathematical Representation of Best Performing Algorithm

Let $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{T} \times \boldsymbol{F}}$represent an input time-series window with Ttime steps and Fsensor features.

The convolutional operation is defined as:

$$h_t = \sigma(W_c * X_t + b_c) \qquad (1)$$

where $W_c$and $b_c$denote convolutional weights and bias, and $\sigma(\cdot)$represents the ReLU activation function.

The BiLSTM layer models temporal dependencies as:

$$h_t = \text{BiLSTM}(h_{t-1}, X_t) \qquad (2)$$

The final prediction is obtained using a softmax classifier:

$$\hat{y} = \text{softmax}(W_o h_T + b_o) \qquad (3)$$

Here, $\hat{y}$ represents the predicted activity class probabilities.

6. Results

Figure 9: Overall Model Performance

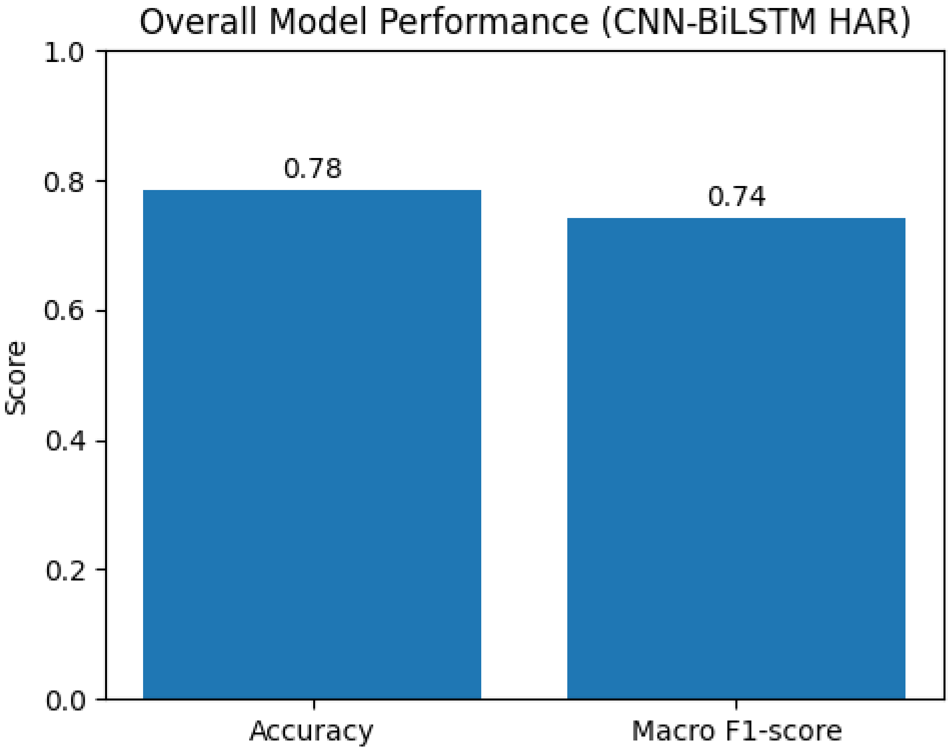
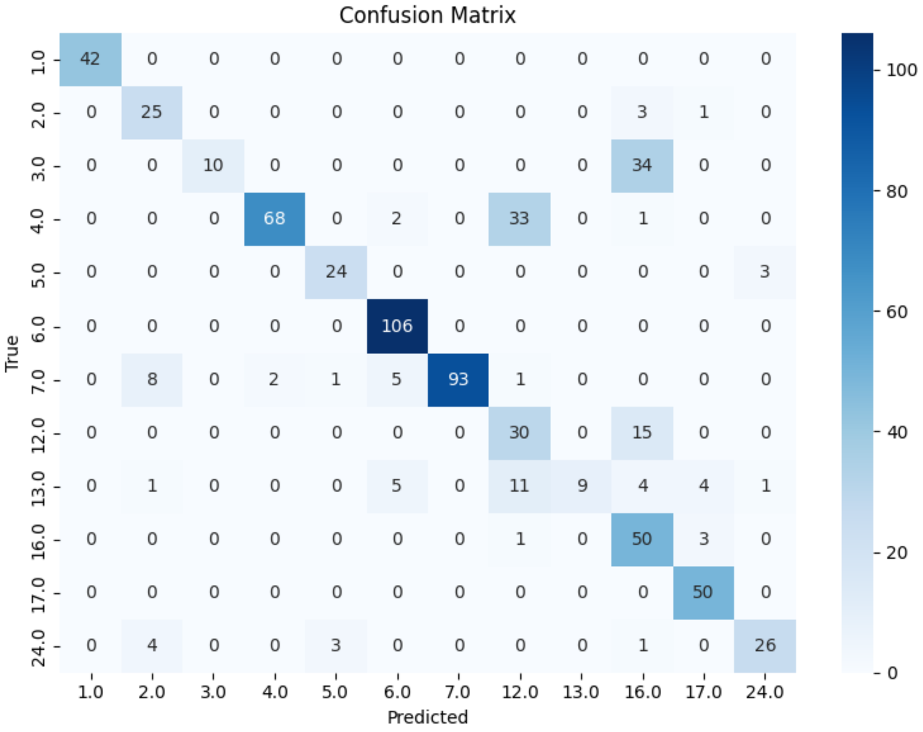


Figure 9: Overall Model Performance in Terms of Accuracy and Macro F1-score.

The CNN-BiLSTM model achieved an overall test accuracy of 78.4% under a subject-wise evaluation protocol. This result demonstrates strong generalization to unseen subjects.

Figure 10: Confusion Matrix



	1.0	2.0	3.0	4.0	5.0	6.0	7.0	12.0	13.0	16.0	17.0	24.0
1.0	42	0	0	0	0	0	0	0	0	0	0	0
2.0	0	25	0	0	0	0	0	0	0	3	1	0
3.0	0	0	10	0	0	0	0	0	0	34	0	0
4.0	0	0	0	68	0	2	0	33	0	1	0	0
5.0	0	0	0	0	24	0	0	0	0	0	0	3
6.0	0	0	0	0	0	106	0	0	0	0	0	0
7.0	0	8	0	2	1	5	93	1	0	0	0	0
12.0	0	0	0	0	0	0	0	30	0	15	0	0
13.0	0	1	0	0	0	5	0	11	9	4	4	1
16.0	0	0	0	0	0	0	0	1	0	50	3	0
17.0	0	0	0	0	0	0	0	0	0	0	50	0
24.0	0	4	0	0	3	0	0	0	0	1	0	26

Figure 10: Confusion Matrix of Activity Classification Results.

The confusion matrix reveals that most activities are correctly classified. However, confusion is observed among motion-similar activities such as standing and walking, as well as stair-related actions. This behavior reflects the inherent difficulty of subject-independent HAR.

Figure 11: Per-Class Precision, Recall, and F1-score

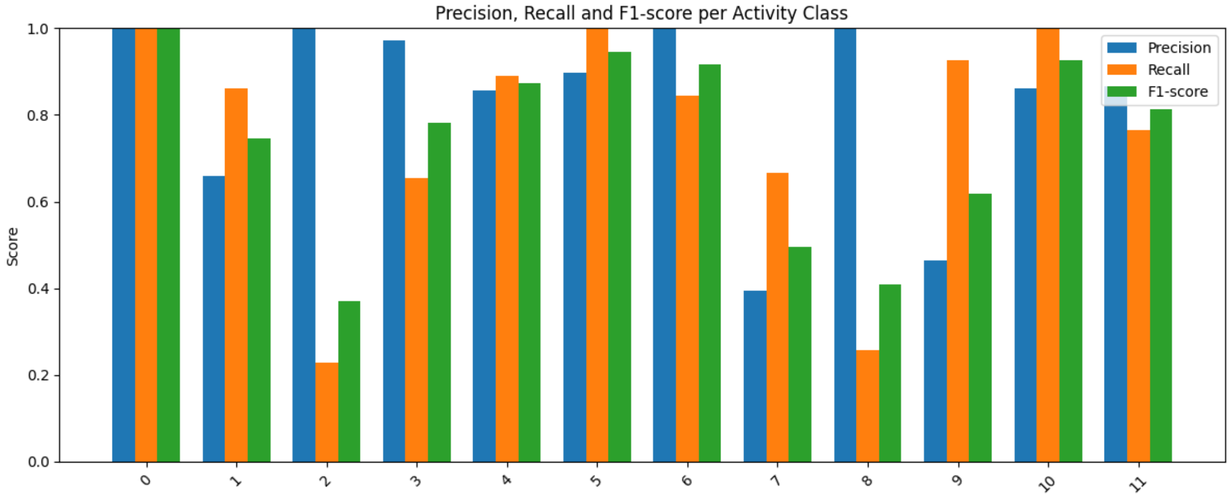


Figure 11: Precision, Recall, and F1-score per Activity Class.

Per-class performance analysis shows high precision and recall for dynamic activities, while static or transitional activities exhibit lower performance due to overlapping motion patterns.

7. Conclusion

This project presented a subject-wise human activity recognition system using wearable sensor data and a CNN–BiLSTM deep learning architecture. A rigorous preprocessing pipeline and subject-independent evaluation strategy were employed to ensure realistic performance. The final model achieved **78.4% accuracy**, demonstrating strong generalization across unseen subjects.

Although high accuracy was achieved for dynamic activities, confusion among motion-similar activities remains a limitation. Future work may explore attention mechanisms, longer temporal windows, and additional sensor modalities to further improve recognition performance.