PART-A: Dataset Description

Dataset Name

MobiFall Dataset v2.0

Dataset Source

- Downloaded from Kaggle
  Link: https://www.kaggle.com/datasets/kmknation/mobifall-dataset-v20

Domain

Human Activity Recognition using Motion Sensor Data

Type of Sensors

This dataset uses **smartphone sensors**:

- **Accelerometer**
- **Gyroscope**

Both record motion in **x, y, z axes**.

Collected Variables (Columns)

Typical file format contains:

Column	Description
timestamp	Time of recorded event
x	Acceleration along x-axis
y	Acceleration along y-axis
z	Acceleration along z-axis

Each `.txt` file contains thousands of rows → one time-series sample.

Activities (Classes Used)

6 classes are selected:

Normal Activities (ADL):

1. **Walking**
2. **Sitting**
3. **Standing**

Fall Activities:

4. **Fall Forward**
5. **Fall Backward**
6. **Fall Sideways**

This makes a **multiclass classification problem**.

Number of Samples

Each subject folder contains multiple activity trials.

Example:

- `sub10/FALLS/BSC_acc_1_2.txt`

Each file ≈ 400–800 rows.

If we choose 10 files per class:
Total ≈ **60 files**

PART-B: Preprocessing Steps

Steps	Why Important?
Parse Timestamps	• Timestamps allow us to understand the temporal sequence of events • Proper datetime format enables time-based operations and analysis • Essential for time series modeling and feature extraction
Chronological Sorting	• Time series analysis requires data in correct temporal order • Many algorithms assume sequential data (e.g., LSTM, sliding windows) • Prevents errors in calculating time-based features (velocity, acceleration)
Handle missing values	• Missing values can cause errors in machine learning models • Gaps in sensor data can lead to incorrect analysis • Proper imputation maintains data continuity for time series
Remove duplicates	• Duplicate rows can bias model training • Outliers/impossible values indicate sensor errors") • Clean data improves model accuracy and reliability
Rename columns	Better readability