Abstract

This report investigates the use of Long Short-Term Memory (LSTM) networks for recognizing human activities and postural transitions based on smartphone accelerometer data. Drawing on the “Smartphone-Based Recognition of Human Activities and Postural Transitions” dataset, two LSTM architectures are explored: one using all 561 available features and another constrained to the top 20 most important features as determined by a RUSBoost-based feature selection procedure. Results show that the **full-feature model** achieves higher accuracy and handles subtle postural transitions more effectively, while the **reduced-feature model** offers faster training but struggles more with underrepresented transitions. Confusion matrices, feature importance graphs, and activity distributions highlight the dataset’s imbalance and the trade-offs between comprehensive input feature sets and computational efficiency.

1. Introduction

Human Activity Recognition (HAR) is critical in digital health applications, supporting accurate tracking of movements and energy expenditure for daily tasks. Beyond standard activities like walking and sitting, postural transitions (e.g., **Sit-to-Stand**, **Stand-to-Lie**) pose additional classification challenges due to their **short duration** and fewer data samples.

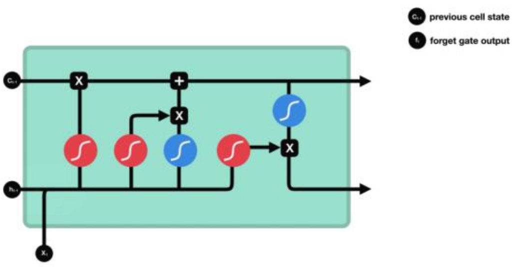
LSTMs are well-suited for this task because they capture **temporal dependencies** in sequential sensor data, preserving information from previous timesteps. This capability enables more robust classification of dynamic and transitional movements than simpler models that primarily rely on static snapshots.

In this work, two **LSTM models** are trained on accelerometer signals from a waist-mounted smartphone, focusing on:

1. **Full-Feature LSTM** (561 features)
2. **Top-20-Feature LSTM** (20 key features from RUSBoost ranking)

Both models are tested on a balanced subset of the dataset, and their performance is measured in terms of accuracy, confusion matrices, and F1 scores. The ultimate goal is to guide decisions regarding computational trade-offs (i.e., training time vs. classification accuracy) in real-world health monitoring scenarios.

2. Brief Overview of LSTM



Illustrated LSTM’s cells

Long Short-Term Memory (LSTM) networks are a specialized form of recurrent neural networks (RNNs) designed to overcome **vanishing and exploding gradient** problems. An LSTM cell typically consists of **input, forget, and output gates**, which control the flow of information. By maintaining a **cell state** across time, LSTMs can selectively remember or forget specific time-step information, making them highly effective for tasks requiring **temporal context**, such as activity recognition and speech processing.

In contrast to simpler RNNs, LSTMs handle longer sequences without losing past information too quickly, enabling them to detect nuanced transitions in activity data. This makes them an appealing choice for modeling sensor sequences that capture human motions and postures across time.

3. Dataset Description and Analysis

3.1 Dataset Origin

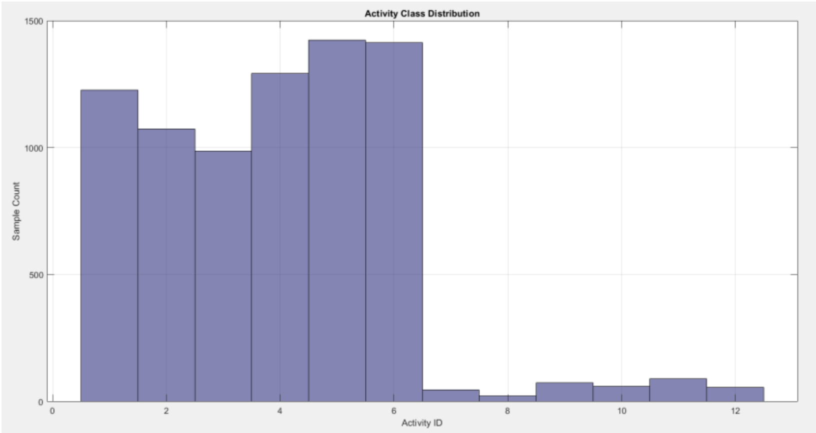
The **"Smartphone-Based Recognition of Human Activities and Postural Transitions"** dataset (UCI repository) contains accelerometer measurements from 30 volunteers performing **12 labeled activities**:

1	**WALKING**
2	**WALKING_UPSTAIRS**
3	**WALKING_DOWNSTAIRS**
4	**SITTING**
5	**STANDING**
6	**LAYING**
7	**STAND_TO_SIT**
8	**SIT_TO_STAND**
9	**SIT_TO_LIE**
10	**LIE_TO_SIT**
11	**STAND_TO_LIE**
12	**LIE_TO_STAND**

Each sample is associated with a **561-dimensional** feature vector, capturing time- and frequency-domain characteristics, as well as statistical and correlation metrics. According to the dataset documentation, features are **already normalized** within $[-1,1]$, so **no additional normalization** was performed.

3.2 Distribution of Activities

Certain activities (e.g., WALKING, SITTING, STANDING) are abundant, each with well over 1,000 samples. Others, such as **STAND_TO_SIT** or **SIT_TO_STAND**, have far fewer entries (fewer than 100 in some cases). This imbalance complicates training, particularly for rarer transitions.



Activity Class Distribution

Notable Imbalances

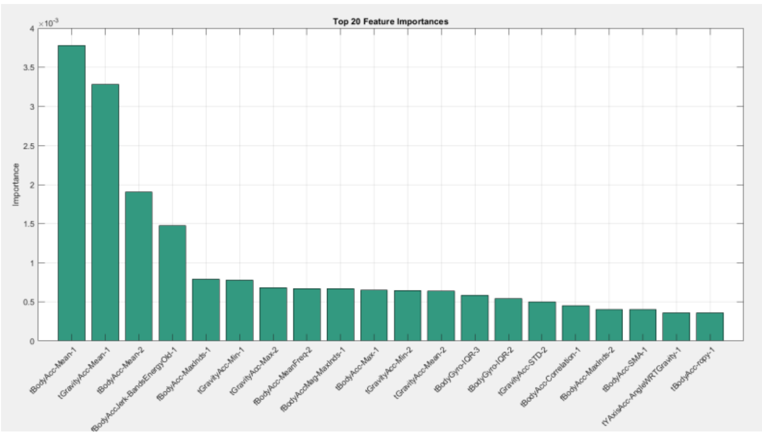
Dynamic (WALKING, UPSTAIRS, DOWNSTAIRS) and **static** (SITTING, STANDING, LAYING) classes each surpass 1,000 samples.

Transitional classes (e.g., SIT_TO_STAND) have fewer than 100 samples each.

Such skew can cause the model to misclassify transitions more frequently if not addressed via techniques like **oversampling** or specialized data augmentation.

3.3 Feature Importance via RUSBoost

RUSBoost a method combining random under-sampling of the majority classes and boosting to improve minority-class recognition was applied to rank all **561 features**. The top 20 features included various time-domain measurements (e.g., **tBodyAcc-Mean-1**, **tGravityAcc-Mean-1**) and frequency-domain measures (e.g., **fBodyAccJerk-BandsEnergyOld-1**).



Top 20 Feature Importances

Key Insights

tBodyAcc-Mean-1 (Feature 1) had the highest relevance (0.0038).

tGravityAcc-Mean-1 (Feature 41) followed (0.0033).

Frequency-based features like fBodyAccJerk contributed substantially.

By restricting the LSTM input to these **top 20 features**, training times decrease. However, the **loss in feature richness** can degrade performance, especially for subtle transitions that require more comprehensive sensor data.

4. Experiment Methodology

4.1 Data Splitting Strategy

The original dataset was split into **training, validation, and test** subsets. To **preserve** minority classes in each subset, **stratified splitting** was employed, ensuring transitional classes were not lost. The test set was further divided in half, producing a **validation** set (50% of test data) and a **final test** set (remaining 50%).

Training Set: Used to learn network weights, with oversampling of minority transition classes.

Validation Set: Monitored performance during training (hyperparameter tuning, early stopping checks).

Test Set: Provided an **unbiased** evaluation of final model performance.

4.2 LSTM Architectures

4.2.1 Full-Feature LSTM

- **Input size**: 561 features per timestep.
- **LSTM layer**: 100 hidden units; OutputMode = 'last'.
- **Dropout layer**: 20% dropout to prevent overfitting.
- **Fully connected layer**: 12 output units, matching the 12 classes.
- **Softmax + classification**: Produces probability distributions and uses cross-entropy loss.

4.2.2 Top-20-Feature LSTM

- **Input size**: 20 features selected via RUSBoost ranking.
- **LSTM layer**: Same configuration (100 hidden units, OutputMode = 'last').
- **Dropout layer**: 20% dropout.
- **Fully connected layer**: 12 output units.

- **Softmax + classification**: Similar probability output layer.

4.3 Training Parameters

- **Optimizer**: Adam, initial learning rate 0.001.
- **Mini-batch size**: 64.
- **Epochs**: 5 (preliminary trials suggested quick convergence by epoch 3–5).
- **Validation frequency**: Every 10 iterations for consistent performance monitoring.
- **Sequence length**: 5 timesteps, balancing coverage of temporal patterns with computational feasibility.

Both models underwent the **same** hyperparameter schedule, except for differences in input dimensions (561 vs. 20). Because the dataset's features are already normalized within [−1,1], **no additional normalization** was introduced.

4.4 Training Time Comparison

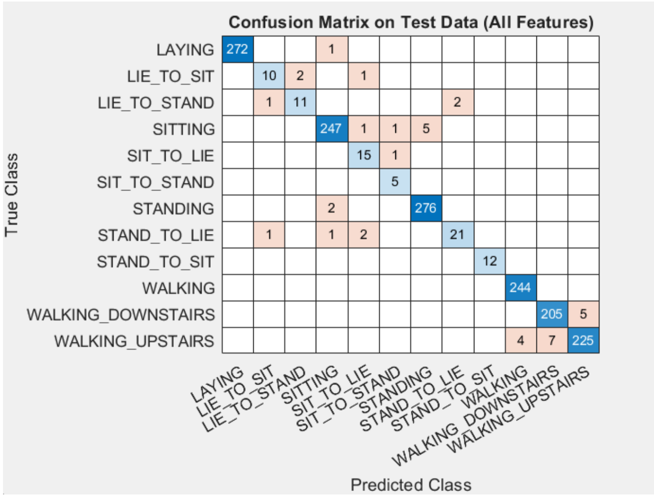
- **Full-Feature LSTM**: ~50 seconds total training time.
- **Top-20-Feature LSTM**: ~33 seconds, roughly **34% faster**.

This speedup can be significant in **real-time** or resource-limited applications.

5. Results and Analysis

5.1 Full-Feature LSTM Model

Class	Precision	Recall	F1_Score
WALKING	0.98387	1.0	0.99187
WALKING_UPSTAIRS	0.97826	0.95339	0.96567
WALKING_DOWNSTAIRS	0.96698	0.97619	0.97156
SITTING	0.98406	0.97244	0.97822
STANDING	0.98221	0.99281	0.98748
LAYING	1.0	0.99634	0.99817
STAND_TO_SIT	1.0	1.0	1.0
SIT_TO_STAND	0.71429	1.0	0.83333
SIT_TO_LIE	0.78947	0.9375	0.85714
LIE_TO_SIT	0.83333	0.76923	0.8
STAND_TO_LIE	0.91304	0.84	0.875
LIE_TO_STAND	0.84615	0.78571	0.81481
Overall	0.97658	0.97658	0.97658



Confusion Matric of the full-feature LSTM

Performance

- **Test Accuracy**: ~97.66%.
- **Static/Dynamic Classes**: Near-perfect F1 scores (e.g., LAYING ~99.82%, STANDING ~98.75%).
- **Transitional Classes**: SIT_TO_STAND and LIE_TO_SIT had lower F1 (~80–83%), indicating greater classification difficulty due to **limited samples** and **short-lived** transitions.

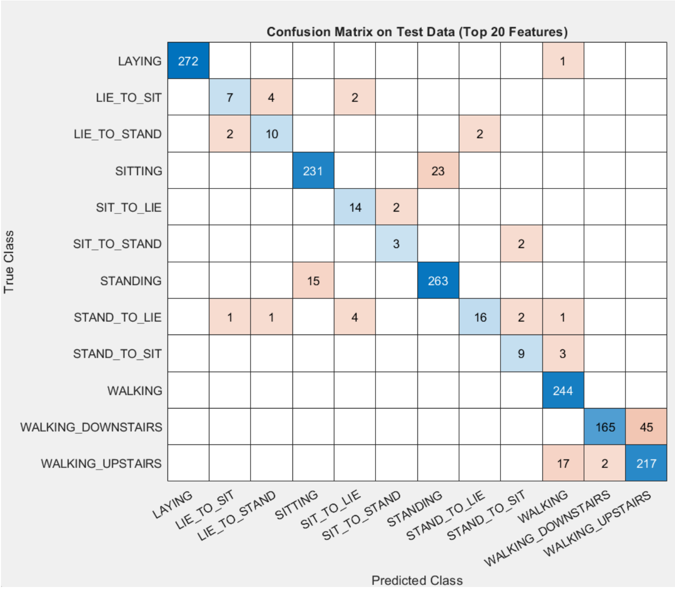
Confusion Matrix Observations

- Generally **strong diagonal** dominance, confirming high accuracy.
- **SITTING vs. STANDING** saw some confusion, reflecting posture similarities.
- **WALKING_DOWNSTAIRS vs. WALKING_UPSTAIRS** also overlapped, likely due to comparable leg movements.

Though oversampling of transitions helped, rare classes still exhibited some misclassifications, underscoring the inherent difficulty in distinguishing subtle shift patterns.

5.2 Top-20-Feature LSTM Model

Class	Precision	Recall	F1_Score
WALKING	0.91729	1.0	0.95686
WALKING_UPSTAIRS	0.82824	0.91949	0.87149
WALKING_DOWNSTAIRS	0.98802	0.78571	0.87533
SITTING	0.93902	0.90945	0.924
STANDING	0.91958	0.94604	0.93262
LAYING	1.0	0.99634	0.99817
STAND_TO_SIT	0.69231	0.75	0.72
SIT_TO_STAND	0.6	0.6	0.6
SIT_TO_LIE	0.7	0.875	0.77778
LIE_TO_SIT	0.7	0.53846	0.6087
STAND_TO_LIE	0.88889	0.64	0.74419
LIE_TO_STAND	0.66667	0.71429	0.68966
Overall	0.91835	0.91835	0.91835



Confusion Matric of the top-20-feature LSTM

Performance

- **Test Accuracy**: ~91.84%.
- **Static Classes**: LAYING remained highly accurate (F1 ~99.82%).
- **Dynamic Classes**: WALKING_DOWNSTAIRS (F1 ~87.53%) was decent but **worse than** the full-feature version.
- **Transitions**: SIT_TO_STAND (F1 ~60%) and LIE_TO_SIT (~60.87%) suffered most, reflecting their reliance on finer sensor details not captured in the reduced feature set.

Confusion Matrix Observations

- **Greater confusion** among SITTING/STANDING and UPSTAIRS/DOWNSTAIRS. E.g., many STANDING samples predicted as SITTING and vice versa.
- Transitional classes accounted for fewer samples but still showed notable errors, such as SIT_TO_LIE misclassified as STAND_TO_LIE.

Implications

While the top-20 model reduced **training time** and computational overhead, it lacks the full sensor nuance. Postural transitions are especially impacted, as they require **rich** feature representations to distinguish subtle movements.

6. Discussion

6.1 Key Findings

1. **Accuracy vs. Efficiency**:
   - The **Full-Feature LSTM** excelled at ~97.66% accuracy, robustly classifying static, dynamic, and transitional classes.
   - The **Top-20-Feature LSTM** reached ~91.84%, offering ~34% faster training but struggling with transitions.
2. **Transition-Class Challenges**:
   Even with oversampling, transitions like SIT_TO_STAND or LIE_TO_SIT remain difficult. Their **lower data availability** and short duration hamper the network's ability to extract stable patterns, particularly when features are limited.
3. **Confusion Hotspots**:
   - **SITTING vs. STANDING** and **UPSTAIRS vs. DOWNSTAIRS** often overlap, reflecting subtle differences in posture or motion direction.
   - Fewer data points for transitions yield higher misclassification rates.

6.2 Practical Trade-Offs

- **Full-Feature Model**:
  - **Pros**: Best overall accuracy, particularly for transitions and highly dynamic classes.
  - **Cons**: Higher computational load (~50 seconds training) and larger memory footprint.
- **Reduced-Feature Model**:
  - **Pros**: Faster (~33 seconds) and lighter, potentially suitable for resource-limited devices.
  - **Cons**: Loses discriminatory power for closely related or infrequent activities.

6.3 Potential Improvements

- **Augmentation of Transitional Classes**:
  Using advanced resampling or synthetic data generation might alleviate underrepresentation, improving the detection of subtle transitions.
- **Adaptive LSTM Architectures**:
  Employing deeper or bidirectional LSTM layers could further refine classification, though at a potential computational cost.
- **Hybrid Feature Selection**:
  Combining domain knowledge (e.g., motion physics) with automatic methods like RUSBoost might better capture relevant features for transitions.

7. Conclusion

This report examined **LSTM-based** recognition of human activities and postural transitions using the UCI "Smartphone-Based Recognition of Human Activities and Postural Transitions" dataset. Two LSTM models were compared:

1. **Full-Feature LSTM (561 features)**
    - Achieved ~97.66% accuracy.
    - Strong performance on static, dynamic, and transitional tasks alike.
    - Required ~50 seconds of training.
2. **Top-20-Feature LSTM**
    - Achieved ~91.84% accuracy.
    - Trained ~34% faster (~33 seconds).
    - Struggled with transitional classes due to reduced sensor information.

Overall, **feature richness** significantly impacts classification particularly for rare transitions that demand finer sensor details. For **maximum accuracy**, especially in clinical or safety-critical applications, the full-feature approach is optimal. Conversely, in **resource-limited** or real-time scenarios, the reduced-feature model might be adequate if a modest accuracy drop is acceptable. In practice, further data augmentation, architectural tuning, or hybrid feature selection could help close this gap, ensuring robust, efficient activity recognition for next-generation health monitoring solutions.