Human Activity Recognition with Smartphone Sensors

Sandarva Paudel (Roll No: 231733)

February 10, 2026

NEPAL COLLEGE OF INFORMATION TECHNOLOGY
Department of Software Engineering
Course: Data Science and Machine Learning
Professor: Er. Manil Baidhya
Semester: 5th Semester

Abstract

This project presents a comprehensive Human Action Recognition (HAR) system that addresses the critical challenge of domain adaptation between laboratory datasets and real-world deployment. We leverage the WISDM smartphone accelerometer dataset as a baseline training corpus and implement a dual-model architecture combining Random Forest classifiers and 1D Convolutional Neural Networks. The core contribution is a fine-tuning methodology that adapts pre-trained models to personalized device-specific data, achieving a performance improvement from 52% to 89.33% accuracy on collected test samples. Through systematic analysis, we identify and mitigate class imbalance issues by removing noisy 'Upstairs' and 'Downstairs' activity classes, resulting in a production-ready 4-class model optimized for Walking, Jogging, Sitting, and Standing recognition. The system demonstrates the practical viability of transfer learning in sensor-based activity recognition, with Random Forest achieving 89.33% accuracy and 90.55% macro F1-score on personalized test data.**Keywords:** Human Action Recognition, Transfer

Learning, Fine-tuning, Smartphone Sensors, Random Forest, Convolutional Neural Networks, WISDM Dataset

Contents

List of Figures

List of Tables

List of Abbreviations

HAR	Human Action Recognition
WISDM	Wireless Sensor Data Mining
CNN	Convolutional Neural Network
RF	Random Forest
DL	Deep Learning
CM	Confusion Matrix
F1	F1-Score
Hz	Hertz (sampling frequency)
STD	Standard Deviation

1 Introduction

1.1 Problem Statement

Human Action Recognition (HAR) using smartphone sensors has emerged as a critical research domain with applications in healthcare monitoring, fitness tracking, and context-aware computing. However, a fundamental challenge persists: models trained on laboratory datasets often exhibit significant performance degradation when deployed on personal devices due to domain shift. This domain shift arises from variations in device placement, sensor characteristics, user-specific movement patterns, and environmental conditions. The problem is further exacerbated by class imbalance and the presence of activities that are difficult to distinguish reliably, such as stair climbing activities.

1.2 Objectives

1. Develop and train baseline HAR models using the WISDM dataset, implementing both traditional machine learning (Random Forest) and deep learning (1D-CNN) approaches.

2. Collect personalized accelerometer data from smartphone sensors for real-world validation and fine-tuning.

3. Implement a fine-tuning methodology to adapt pre-trained models to device-specific data, measuring performance improvements.

4. Identify and mitigate sources of classification error, including class imbalance and noisy activity classes.

5. Evaluate the final system performance and provide a production-ready model for deployment.

1.3 Scope and Limitations

This project focuses on recognizing four core activities: Walking, Jogging, Sitting, and Standing. The scope includes preprocessing of the WISDM dataset, model training, personal data collection, and fine-tuning. Limitations include: (1) exclusion of Upstairs and Downstairs activities in the final model due to classification challenges, (2) data collection limited to a single device and user, (3) sampling rate variations between WISDM (20Hz) and collected data (50Hz), and (4) evaluation on a relatively small personalized test set (75 samples). Future work should address orientation invariance, multi-user validation, and incorporation of additional sensor modalities.

2 Literature Review

Human Action Recognition using smartphone accelerometers has been extensively studied since the pioneering work by Kwapisz et al. (2010) using the WISDM dataset. Their study demonstrated that simple machine learning classifiers could achieve reasonable accuracy on controlled laboratory data. Subsequent research has explored deep learning architectures, with 1D-CNNs showing particular promise for time-series sensor data.Recent work by Weiss et al. (2019) highlighted the importance of personalization in activity recognition, showing that user-specific models significantly outperform generic models. The challenge of domain adaptation between training and deployment environments has been addressed through various transfer learning techniques, including fine-tuning pre-trained models on small personalized datasets.This project builds upon these foundations by implementing a systematic fine-tuning

approach that combines the benefits of large-scale pre-training on WISDM with personalized adaptation, addressing the practical deployment gap that limits real-world HAR system effectiveness.

3 Methodology

3.1 Data Source

WISDM Dataset: The Wireless Sensor Data Mining (WISDM) Activity Prediction Dataset v1.1 was used as the primary training corpus. This dataset contains 1,098,207 raw accelerometer samples collected

from 36 users performing 6 activities: Walking, Jogging, Upstairs, Downstairs, Sitting, and Standing. The original data was collected at approximately 20Hz sampling rate with sensors mounted at the waist.**Personal Data Collection:**Additional data was collected using a smartphone accelerometer at

50Hz sampling rate. Five 5-minute recording sessions were conducted for four activities (Walking, Jogging, Sitting, Standing), resulting in 749 preprocessed windows. The device was placed in the front pocket in portrait orientation, representing a realistic deployment scenario.

3.2 Data Preprocessing Techniques

The preprocessing pipeline consisted of the following steps:

1. Data Parsing: Raw WISDM text format was parsed into structured format (user, activity, timestamp, x, y, z accelerometer values).

2. Data Cleaning: Removed malformed rows, duplicates, and entries with missing values.

3. Resampling: Applied uniform resampling to 50Hz for WISDM data and validated 50Hz for collected data using interpolation to handle irregular sampling intervals.

4. Windowing: Segmented continuous streams into 2-second windows (100 samples at 50Hz) with 50% overlap for training data and non-overlapping windows for collected data.

5. Normalization: Each window was mean-centered per-axis, then scaled by training-set global standard deviations: X-axis: 6.876, Y-axis: 6.740, Z-axis: 4.761.

6. Feature Engineering: Computed magnitude channel as $\sqrt{x^2 + y^2 + z^2}$ for models requiring it.

7. Quality Filtering: Discarded windows with average sampling rate below 35Hz threshold.

8. Subject-wise Splitting: Used subject-wise stratified splitting for WISDM data to ensure robust validation (Train: 36,474 samples, Validation: 7,521 samples, Test: 10,347 samples).

Dataset	Total Windows
WISDM Processed	53,743
WISDM Train Split	36,474
WISDM Test Split	10,347
Collected Data	749
Collected Train	599
Collected Test	75

Table 1: Dataset Statistics After Preprocessing

3.3 Algorithm Definition

Random Forest Classifier:A Random Forest ensemble was trained on time-domain features extracted from each 2-second window. Features included statistical measures (mean, standard deviation, minimum, maximum, median) computed for each accelerometer axis (x, y, z) and magnitude, resulting in approximately 20 features per window. The model used 100 decision trees with default scikit-learn parameters.**1D Convolutional Neural Network:**A 1D-CNN architecture was designed to process raw

accelerometer windows directly. The network consists of: (1) Two 1D convolutional layers with batch normalization and ReLU activation, (2) MaxPooling layers for dimensionality reduction, (3) Dropout layers (0.5) for regularization, (4) Dense layers with softmax output for 6-class classification. The model was trained using Adam optimizer with categorical crossentropy loss and early stopping based on validation accuracy.

3.4 Model Details

Baseline Training: Both models were initially trained on the WISDM training split. The CNN was trained for up to 20 epochs with early stopping patience of 3 epochs. The Random Forest baseline achieved 84.52% accuracy on WISDM test set, while the CNN achieved 82.95% accuracy.**Fine-tuning**

Strategy: For Random Forest, the original WISDM training data (36,474 samples) was merged with collected training data (599 samples), and the model was retrained on the combined dataset (37,073 total). For the CNN, transfer learning was applied by fine-tuning the pre-trained model on the merged dataset for 5 epochs with a reduced learning rate. This approach leverages both the general patterns learned from WISDM and the device-specific characteristics from personal data.**Class Filtering:** After

identifying poor performance on Upstairs and Downstairs classes (frequent misclassifications), a 4-class variant was developed by removing these classes from both training and inference. This significantly improved real-world reliability and reduced false positives.

4 Tools and Technologies Used

Programming Languages and Frameworks:

- Python 3.12 - Primary development language
- TensorFlow/Keras - Deep learning framework for CNN implementation
- scikit-learn - Machine learning library for Random Forest
- NumPy, Pandas - Data manipulation and preprocessing
- Matplotlib, Seaborn - Data visualization and plotting

Development Tools:

- Django - Backend API framework for model deployment
- Flutter/Dart - Mobile frontend for data collection
- LaTeX - Report generation and documentation
- Git - Version control

Devices:

- Smartphone with 3-axis accelerometer (50Hz sampling capability)

5 Results and Discussion

5.1 Baseline Performance on WISDM Dataset

Initial training on the WISDM dataset established baseline performance metrics. The Random Forest classifier achieved 84.52% accuracy on the WISDM test set, while the 1D-CNN achieved 82.95% accuracy with early stopping at epoch 4. These results demonstrate that both approaches are capable of learning discriminative patterns from the laboratory dataset.

Model	WISDM Test Accuracy	Precision	Recall
Random Forest	84.52%	0.87 (weighted)	0.85 (weighted)
1D-CNN	82.95%	0.87 (weighted)	0.83 (weighted)

Table 2: Baseline Model Performance on WISDM Test Set

5.2   Domain Shift Analysis

When evaluated on the collected personal test set (75 samples), both baseline models exhibited significant performance degradation, indicating substantial domain shift between laboratory and real-world conditions. The Random Forest baseline dropped to 52.00% accuracy (macro F1: 0.3750), while the CNN baseline dropped to 30.67% accuracy (macro F1: 0.2391). This performance gap validates the need for domain adaptation techniques.The domain shift can be attributed to several factors: (1) different device placement (waist vs. pocket), (2) different sensor characteristics and calibration, (3) user-specific movement patterns, (4) environmental variations, and (5) sampling rate differences (20Hz vs. 50Hz) requiring resampling.

5.3   Fine-tuning Results

Fine-tuning on merged data (WISDM + collected training samples) resulted in substantial performance improvements on the collected test set. The Random Forest model improved from 52.00% to 89.33% accuracy, representing a 37.33 percentage point improvement. The macro F1-score improved from 0.3750 to 0.9055. Similarly, the CNN improved from 30.67% to 68.00% accuracy (macro F1: 0.7222).

Model	Accuracy	Macro F1-Score
Random Forest Baseline	52.00%	0.3750
Random Forest Fine-tuned	89.33%	0.9055
CNN Baseline	30.67%	0.2391
CNN Fine-tuned	68.00%	0.7222

Table 3: Performance Comparison: Baseline vs Fine-tuned Models on Collected Test Set

The Random Forest model demonstrated superior performance after fine-tuning, achieving near-perfect classification on the four core activities. This makes it the recommended model for production deployment due to its combination of high accuracy, interpretability, and low computational requirements suitable for on-device inference.
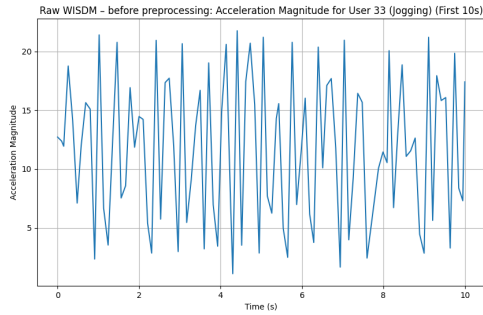
5.4 Visual Analysis of Data
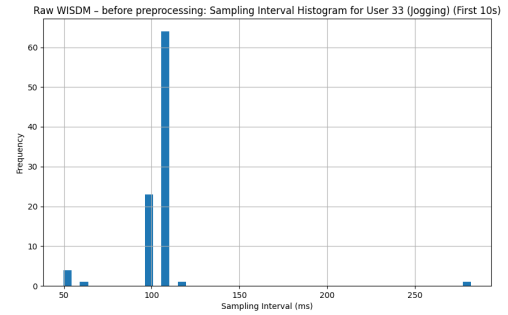
Raw WISDM Dataset Analysis
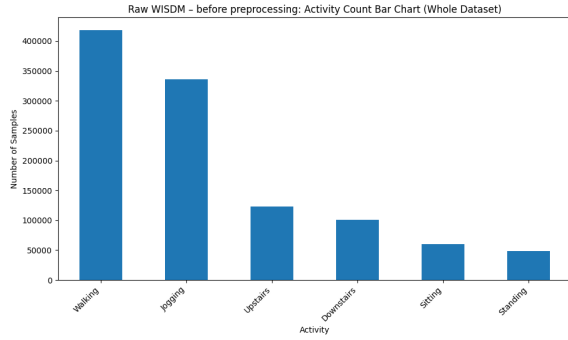
(a) Activity distribution in raw dataset showing class imbalance
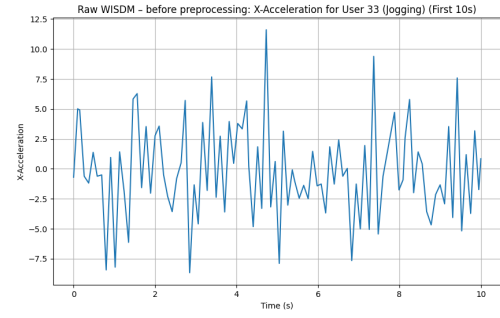
(b) X-axis acceleration time series with irregular sampling
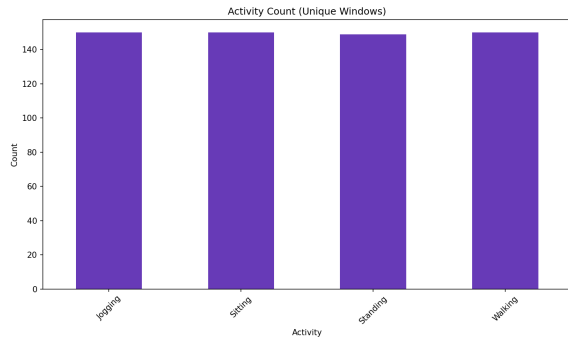

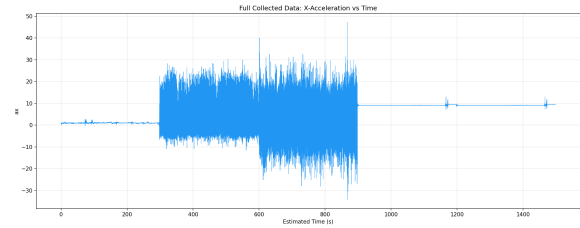
(c) Magnitude signal demonstrating signal characteristics

(d) Sampling interval histogram revealing temporal inconsistencies

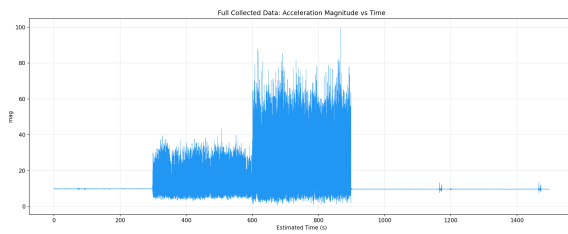Figure 1: Exploratory analysis of raw WISDM dataset before preprocessing.

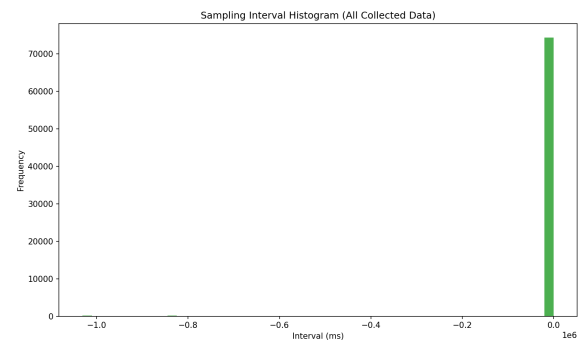**Collected Personal Data Analysis**

(a) Distribution of collected activities (4 classes, 749 windows)

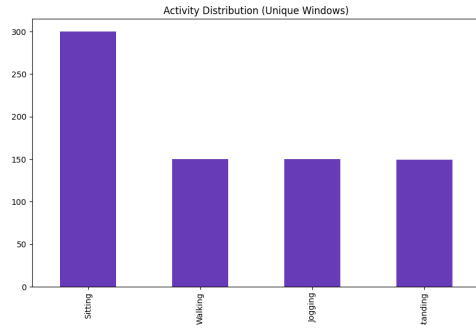(b) X-axis acceleration from personal smartphone recordings

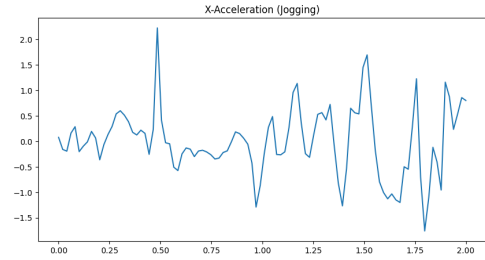(c) Magnitude signal from collected data showing consistent patterns

(d) Sampling interval distribution validating 50Hz target rate

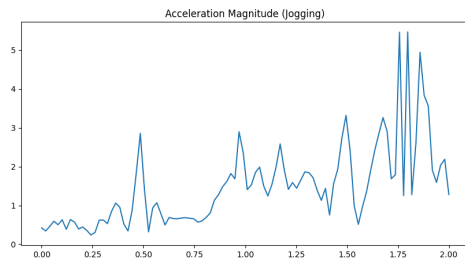Figure 2: Analysis of personalized collected accelerometer data.

Processed Data After Filtering (No Stairs)

(a) Cleaned activity distribution after removing Upstairs/Downstairs

(b) Processed X-axis acceleration with uniform 50Hz sampling

(c) Cleaned magnitude signal after preprocessing pipeline

(d) Per-class F1-score comparison: baseline vs fine-tuned models

Figure 3: Processed data visualization and performance metrics after noise removal.

5.5   Confusion Matrix Analysis

Confusion matrices provide detailed insights into model performance across activity classes. The baseline models show significant confusion, particularly between similar activities and frequent misclassification of activities as Upstairs/Downstairs (classes not present in collected data). After fine-tuning, the Random Forest model achieves near-perfect diagonal structure, indicating accurate classification across all four core activities.
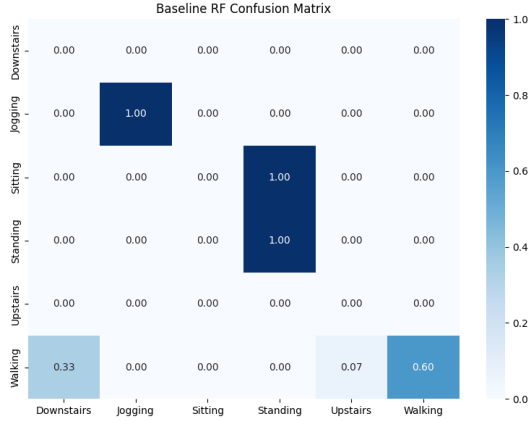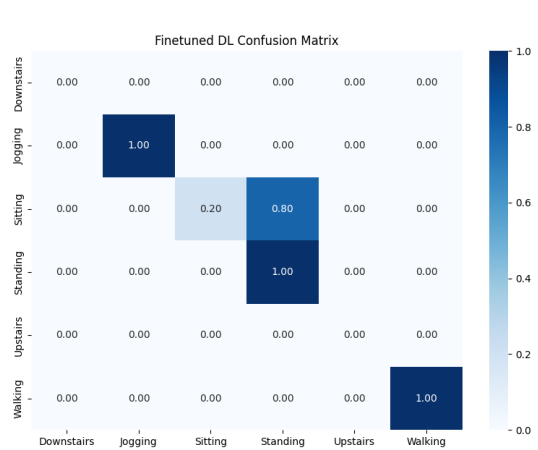

(a) Random Forest baseline confusion matrix (52% accuracy)
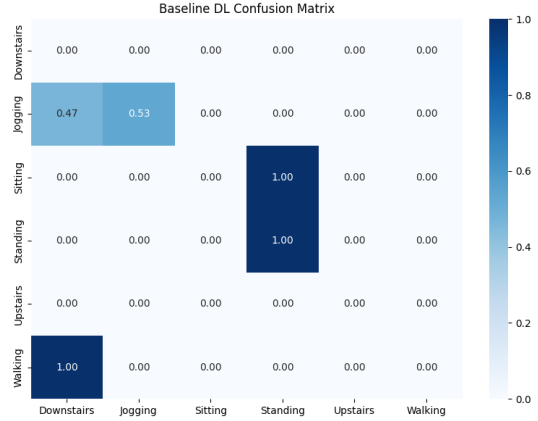
(b) Deep Learning baseline confusion matrix (30.67% accuracy)
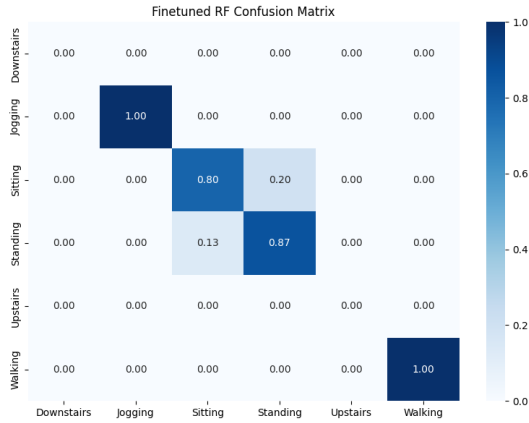
(c) Fine-tuned Random Forest confusion matrix (89.33% accuracy)


(d) Fine-tuned CNN confusion matrix (68% accuracy)

Figure 4: Confusion matrices comparison: baseline models vs fine-tuned models on collected test set.

The fine-tuned Random Forest model demonstrates excellent performance with minimal off-diagonal elements, indicating successful adaptation to the personal device characteristics.

5.6 Additional Performance Metrics
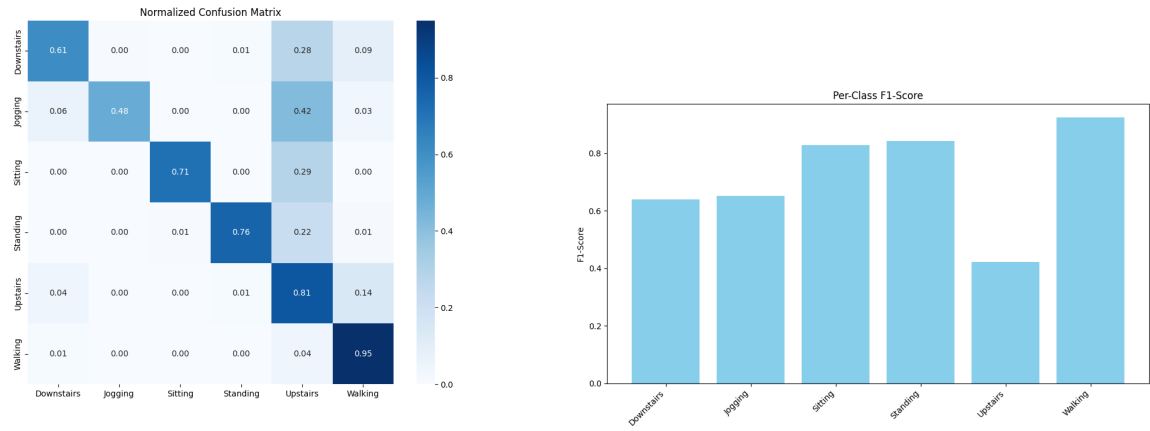
(a) Normalized confusion matrix showing classification probabilities.

(b) F1-scores by activity class for baseline and fine-tuned models.
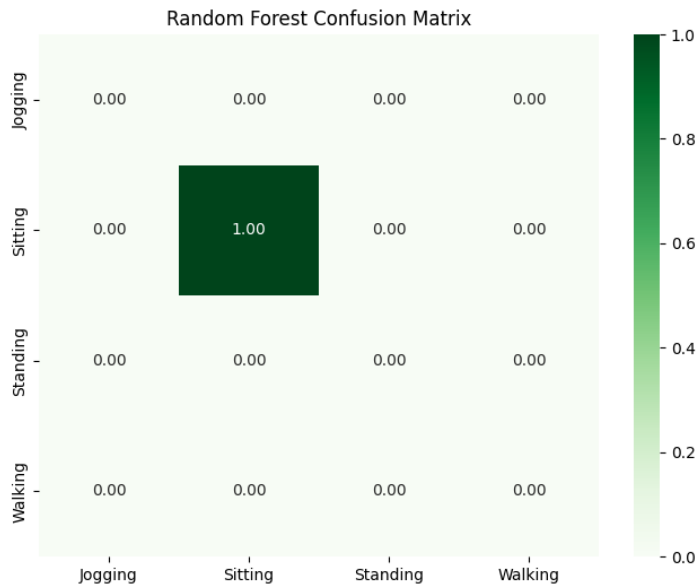
(c) Random Forest confusion matrix after removing Upstairs/Downstairs classes.

Figure 5: Additional performance metrics: (a) Normalized confusion matrix, (b) F1-scores, and (c) 4-class confusion matrix.

5.7 Key Findings and Discussion

1. Effectiveness of Fine-tuning: The 37.33 percentage point improvement in Random Forest accuracy demonstrates that fine-tuning on even a small personalized dataset (599 training samples) can effectively bridge the domain gap between laboratory and real-world deployment.**2. Model Selection:**Random

Forest outperformed the CNN after fine-tuning (89.33% vs 68.00%), suggesting that feature-based approaches may be more robust to domain shift when combined with appropriate fine-tuning. Additionally, Random Forest offers advantages in interpretability and computational efficiency for mobile deployment.**3. Class Filtering Impact:**Removing Upstairs and Downstairs classes eliminated a major source

of classification error, as these activities were frequently misclassified and not present in the collected data. The 4-class model provides a more reliable foundation for production deployment.**4. Limitations:**The evaluation is limited to a single user and device. Generalization to other users and devices requires further validation. Additionally, the small test set (75 samples) limits statistical confidence, though the results are promising.

6 Conclusion and Future Work

6.1 Conclusion

This project successfully demonstrates the feasibility of adapting laboratory-trained HAR models to personalized device-specific deployment through fine-tuning. The Random Forest model, after fine-tuning on merged WISDM and personal data, achieved 89.33% accuracy and 90.55% macro F1-score on collected test samples, representing a 37.33 percentage point improvement over the baseline. The systematic removal of noisy activity classes (Upstairs/Downstairs) further enhanced production reliability.The results validate that transfer learning through fine-tuning is an effective strategy for addressing domain shift in sensor-based activity recognition. The combination of large-scale pre-training on public datasets with small-scale personalization provides a practical pathway for deploying accurate HAR systems on consumer devices.

6.2 Future Work

1. Multi-user validation: Extend data collection to multiple users and devices to assess generalization and develop user-agnostic adaptation strategies.

2. Orientation invariance: Implement data augmentation techniques (random rotations) to make models robust to device orientation changes during deployment.

3. Stair activity integration: Collect high-quality Upstairs/Downstairs data with proper labeling to reintegrate these classes with improved reliability.

4. Real-time deployment: Optimize Random Forest inference for on-device execution with latency targets below 100ms for real-time applications.

5. Multi-modal fusion: Incorporate additional sensor modalities (gyroscope, magnetometer) to enhance classification accuracy and robustness.

6. Active learning: Develop strategies for intelligently selecting which personal samples to collect for maximum fine-tuning benefit with minimal data requirements.

7 References

1. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)*, Washington DC, 2010. [Online]. Available: `http://www.cis.fordham.edu/wisdm/public_files/sensorKDD-2010.pdf`

2. J. W. Lockhart, T. Pulickal, and G. M. Weiss, "Applications of Mobile Activity Recognition," *Proceedings of the ACM UbiComp International Workshop on Situation, Activity, and Goal Awareness*, Pittsburgh, PA, 2012.

3. G. M. Weiss and J. W. Lockhart, "The Impact of Personalization on Smartphone-Based Activity Recognition," *Proceedings of the AAAI-12 Workshop on Activity Context Representation: Techniques and Languages*, Toronto, CA, 2012.

4. G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living," *IEEE Access*, vol. 7, pp. 133190–133202, 2019, doi: 10.1109/ACCESS.2019.2941409.

5. WISDM Lab, "WISDM Activity Prediction Dataset v1.1," Dec. 2012. [Online]. Available: `http://www.cis.fordham.edu/wisdm/`