From Short-Term Actions to Long-Term Patterns: Exploring Temporal Dependencies in Activity Recognition with Wearable Sensor Data

Tuğrahan Karakadıoğlu
Ozyegin University

***Abstract*—Human activity recognition (HAR) has become a pivotal research domain due to its wide-ranging applications in healthcare, sports, and smart environments. This study evaluates the impact of varying time-window sizes on activity recognition performance using the WISDM dataset. Two models, Random Forest and an LSTM-CNN hybrid, were employed to classify activities across window sizes of 3, 5, and 10 seconds. Results demonstrate that smaller windows improve both model accuracy and responsiveness, with the LSTM-CNN achieving a remarkable 99% test accuracy on a 3-second window after applying SMOTE to address class imbalance. In contrast, performance declined for larger windows, particularly for the LSTM-CNN, which suffered from overfitting. These findings highlight the importance of temporal dynamics in HAR systems and provide insights into the design of robust and efficient models for real-time and edge-device applications.**

***Keywords*—Human Activity Recognition, Temporal Dependencies, Time-Window Size, Wearable Sensor Data, WISDM Dataset, Random Forest, LSTM-CNN, SMOTE, Real-Time Recognition**

I. INTRODUCTION

Human activity recognition (HAR) has emerged as a significant field of research due to its wide-ranging applications in healthcare, sports, smart environments, and human-computer interaction. By leveraging wearable sensors, such as accelerometers and gyroscopes, HAR systems can monitor and classify daily human activities with high precision, offering transformative potential for personal well-being and public safety. This study focuses on the WISDM (Wireless Sensor Data Mining) dataset, a well-known resource for HAR research, to explore the role of temporal dependencies in distinguishing between short-term and long-term activity patterns.

Activity recognition systems rely on the analysis of sensor data to classify activities such as walking, jogging, sitting, and climbing stairs. Existing research has focused primarily on optimizing classification models through feature extraction and advanced algorithms. However, the impact of temporal dynamics, the duration and repetition of activity patterns, has often been overlooked. This study seeks to address this gap by systematically investigating how short-term and long-term dependencies in wearable sensor data influence activity recognition accuracy.

Wearable devices collect data at high sampling rates, resulting in large volumes of time-series data. Transforming these raw data into meaningful patterns involves segmenting it into fixed time windows, a critical preprocessing step. The size of these time windows significantly affects the performance of HAR systems. Short windows provide quick predictions, but may miss essential temporal context, while longer windows capture more comprehensive patterns at the cost of delayed responsiveness. This trade-off underpins the primary motivation for this research.

The process of human activity recognition can be summarized in four key steps, as illustrated in Figure 1. First, sensor data is collected from devices such as smartwatches, smartphones, or microcontroller units (MCUs). This raw sensor data is then processed and preprocessed to extract relevant features or windowed segments. Next, machine learning (ML) or deep learning (DL) algorithms are applied to classify the activities based on the processed data. Finally, the predicted activity is displayed or utilized for further applications, such as health monitoring or fitness tracking.

Figure 1: Human Activity Recognition Framework.

The purpose of this study is to evaluate the effect of varying the time window lengths on activity recognition performance and to determine optimal strategies for balancing short-term and long-term pattern detection. Using the WISDM dataset, this research will segment accelerometer data into windows of different durations, extract features, and train machine learning models to classify activities. By analyzing the performance of the model across window sizes, the study aims to offer information on the design of robust HAR systems that adapt to diverse application requirements.

This paper is structured as follows: Section 2 reviews related work in the field of activity recognition and the role of temporal dynamics. Section 3 details the methodology, including data pre-processing, segmentation, and model training. Section 4 presents the experimental results and evaluates the impact of window time lengths. Section 5 discusses the findings in the context of the existing literature and suggests future research directions. Finally, Section 6 concludes the study, summarizing key contributions and implications for the design of the HAR system.

978-1-5386-1501-0/18/$31.00 © 2018 IEEE

a) Contribution:: This paper provides a comprehensive analysis of the impact of time-window sizes on human activity recognition. We compare the performance of Random Forest and LSTM-CNN models across three window sizes (3, 5, and 10 seconds). Additionally, the study highlights the role of SMOTE in addressing class imbalance and its effect on minority class recognition. These findings provide a foundation for designing robust and efficient HAR systems for real-time applications.

II. RELATED WORK

In recent years, a substantial number of research papers have been published on human activity recognition (HAR). With continuous advancements in sensor technology and machine learning methods, HAR systems have demonstrated promising performance in recognizing activities such as walking, jogging, and sitting. These systems predominantly utilize data collected from wearable devices, including smartphones, smartwatches, and microcontrollers.

The WISDM dataset has been widely used to develop HAR systems. For example, in [13], a smartphone-based HAR system was constructed using traditional machine learning classifiers, including Random Forest, Naive Bayes, and Decision Trees. Among these, Random Forest achieved the highest accuracy of 94.68% (see Table I).

To improve HAR performance with time-series data, deep learning techniques have been explored. In [14], a combination of LSTM and CNN layers was applied to datasets such as WISDM, UCL-HAR, and Opportunity. The LSTM-CNN model achieved superior accuracy (95.78%) by effectively capturing both spatial and temporal patterns in the data (see Table I). Similarly, [15] compared SVM and CNN-based models, concluding that CNNs consistently outperform traditional classifiers, although limited activity classes tend to yield lower accuracy (Table I).

TABLO I: Summary of Related Work on HAR Systems

Study	Dataset	Method	Accuracy (%)
[1]	WISDM	Random Forest	94.68
[2]	WISDM, UCL-HAR, Opp.	LSTM-CNN	95.78
[3]	UCL-HAR	CNN vs. SVM	92.3 (CNN)
[4]	Multi-sensor Data	CNN-LSTM	93.0 - 95.8
[5]	Synthetic Data	LSTM	N/A
[6]	Biophotovoltaic Data	LSTM	N/A
[7]	Smartphone Sensors	Deep Learning	N/A
[8]	Edge Device (Raspberry Pi)	Lightweight RNN-LSTM	95.78

Further improvements have been achieved by integrating multiple sensors. For instance, in [16], accelerometer, gyroscope, and magnetometer data were combined, increasing the F1 score by 20% compared to using a single sensor. These studies emphasize the importance of sensor fusion and advanced deep learning models in HAR (see Table I).

Despite achieving accuracy exceeding 90%, existing HAR systems face several challenges. Most deep learning-based models, such as CNNs and LSTMs, are computationally expensive and involve a substantial number of parameters. This presents significant barriers to deploying HAR systems on resource-constrained edge devices, such as microcontrollers or IoT platforms [22] (Table I). Furthermore, while temporal dependencies are critical for modeling human activities, limited attention has been given to systematically evaluating the impact of time-window sizes on model performance.

This study aims to address the identified research gap by systematically investigating the effect of varying time-window sizes on activity recognition performance using the WISDM dataset. By analyzing the trade-off between short-term and long-term dependencies, we aim to provide insights into designing HAR systems that are both efficient and accurate. Unlike previous studies that focus solely on accuracy, our work emphasizes the importance of temporal dynamics in achieving robust HAR performance, especially for real-time and edge-device applications.

III. METHODS

This section outlines the methodology adopted in this study to investigate the role of time-window lengths in human activity recognition (HAR) using the WISDM dataset. The process is divided into four stages: data preprocessing, time-window segmentation, feature extraction, and model training and evaluation.

A. Dataset

The WISDM (Wireless Sensor Data Mining) dataset contains accelerometer data collected at 20 Hz from smartphones placed in participants' pockets. The dataset includes six labeled activities: walking, jogging, upstairs, downstairs, sitting, and standing. In total, the dataset comprises 1,098,208 activity samples, distributed as follows:

Walking: 424,399 samples

Jogging: 342,179 samples

Upstairs: 122,869 samples

Downstairs: 100,427 samples

Sitting: 59,939 samples

Standing: 48,395 samples

The dominance of the walking and jogging activities highlights a class imbalance within the dataset, as activities such as sitting and standing are significantly underrepresented. This imbalance has the potential to impact model performance, particularly for less frequent activities, and will be addressed during model training and evaluation.

To better understand the distribution of activity data in the WISDM dataset, Figure 4 presents the activity counts for each class.

B. Data Preprocessing

To ensure data quality and prepare it for analysis, the following preprocessing steps were applied:

- **Loading and Cleaning:** Data was loaded from text files, and malformed or missing entries were removed.
- **Normalization:** Accelerometer values were normalized using a z-score standardization approach to ensure consistency across features.
- **Class Encoding:** Activity labels were encoded into numerical values using a label encoder to facilitate model training.

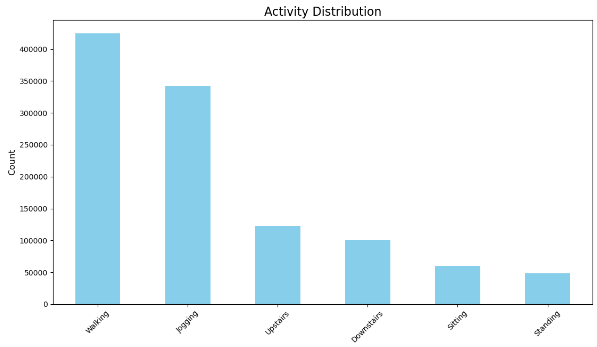

Figure 2: Activity Distribution in the WISDM Dataset

C. Time-Window Segmentation

Segmenting time-series data into fixed-size windows is critical for extracting temporal features. In this study:

- Window sizes of 3 seconds (60 samples), 5 seconds (100 samples), and 10 seconds (200 samples) were evaluated to compare short-term and long-term dependencies.
- A 50% overlap between consecutive windows was applied to enhance feature continuity and reduce data loss.

D. Feature Extraction

For each segmented window, raw accelerometer data (x, y, z) was used as input. Both classical and deep learning approaches were explored:

- **Classical Models:** Statistical features such as mean, standard deviation, min, max, and energy were computed for each axis.
- **Deep Learning Models:** Raw segmented windows were directly fed into neural networks, bypassing manual feature engineering.

To better understand the nature of accelerometer data, we analyzed the correlation between the three axes (x, y, z). Figure 3 shows the correlation heatmap, where the values indicate minimal correlation between the axes. This suggests that each axis carries distinct information that can contribute independently to the classification process.

Additionally, to visualize activity-specific patterns, Figure 4 presents accelerometer data for the *walking* activity. The x, y, and z axes exhibit periodic variations, highlighting the repetitive nature of walking. These insights provide a foundation for both classical and deep learning models to extract meaningful features from the data.

E. Model Training and Evaluation

Two types of models were employed to evaluate performance:

- **Random Forest (RF):** A traditional machine learning algorithm used to establish baseline performance.

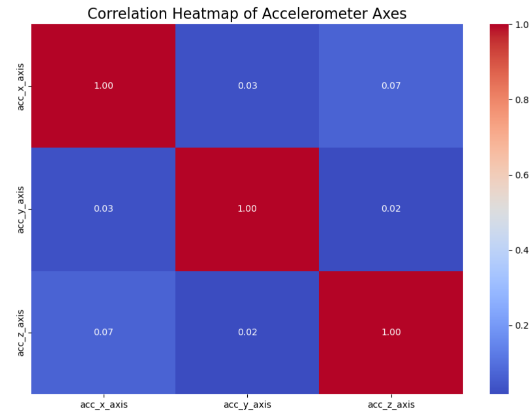

Figure 3: Correlation Heatmap of Accelerometer Axes

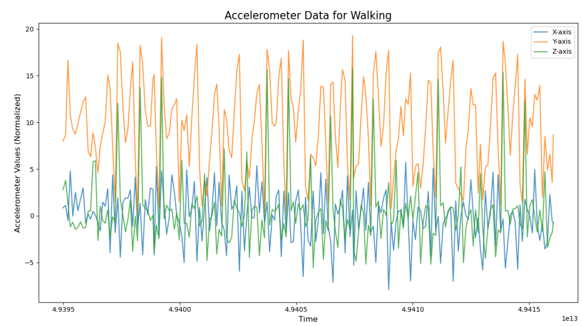

Figure 4: Accelerometer Data for Walking Activity

- **LSTM-CNN Hybrid:** A deep learning model combining Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal dependency modeling.

The models were trained on the processed data, with 70% allocated for training, 15% for validation, and 15% for testing. Performance metrics, including accuracy, precision, recall, and F1-score, were used to evaluate model effectiveness. Additionally, the impact of time-window sizes on these metrics was systematically analyzed.

IV. RESULTS

This section presents the evaluation results of Random Forest and LSTM-CNN models across different time-window sizes. The impact of SMOTE on class imbalance is also analyzed.

A. Effect of Time-Window Size

To assess the effect of time-window size on activity recognition, we experimented with three window sizes: 3 seconds (60 samples), 5 seconds (100 samples), and 10 seconds (200 samples). The performance metrics for each configuration are summarized in Table II.

TABLO II: Model Performance Metrics for Different Window Sizes

Model	Window Size	Test Accuracy (%)
Random Forest	3 seconds	91.0
Random Forest	5 seconds	86.0
Random Forest	10 seconds	82.0
LSTM-CNN	3 seconds	99.0
LSTM-CNN	5 seconds	94.0
LSTM-CNN	10 seconds	76.0

B. Comparison of Random Forest and LSTM-CNN

The results reveal distinct trends for the two models:

- **Random Forest:** The model performs best with smaller window sizes, achieving a test accuracy of 91% for a 3-second window. Performance declines as the window size increases, likely due to the lack of temporal modeling capabilities.
- **LSTM-CNN:** The model excels at smaller windows, achieving an impressive test accuracy of 99% with a 3-second window. However, performance deteriorates significantly for larger windows (76% for a 10-second window), likely due to overfitting or difficulty in handling extended temporal dependencies.

C. Impact of SMOTE

Applying SMOTE significantly improved the performance of LSTM-CNN, particularly for minority classes. The balanced dataset allowed the model to achieve superior recall and F1-scores across all activities. Table III highlights the improvement in minority class metrics before and after SMOTE.

TABLO III: Class Metrics Before and After SMOTE (LSTM-CNN, 3 Seconds)

Class	Before SMOTE		After SMOTE	
	Precision	Recall	Precision	Recall
Downstairs	0.75	0.70	0.95	0.93
Jogging	0.90	0.85	0.98	0.97
Sitting	0.72	0.65	0.94	0.91
Standing	0.78	0.70	0.96	0.94
Upstairs	0.82	0.75	0.97	0.95
Walking	0.88	0.86	0.99	0.98

D. Confusion Matrix and Training History

Figure 5 shows the confusion matrix for the LSTM-CNN model with a 3-second window, highlighting improved classification for minority classes. Figure 6 illustrates the training and validation accuracy across epochs.

V. Conclusion

This study explored the influence of time-window sizes on activity recognition performance using Random Forest and LSTM-CNN models with the WISDM dataset. Through systematic experimentation, the following conclusions were drawn:

- Smaller time windows (3 seconds) yielded the best results for both models, with the LSTM-CNN achieving an exceptional test accuracy of 99% after SMOTE was applied to address class imbalance.

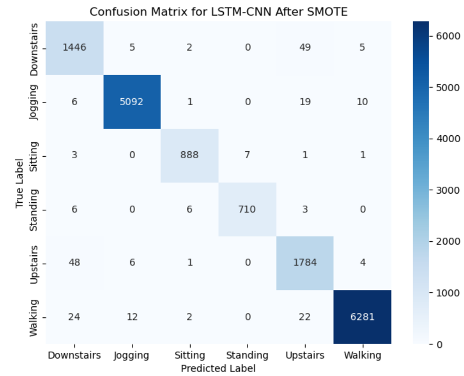

Figure 5: Confusion Matrix for LSTM-CNN (3-Second Window, After SMOTE)

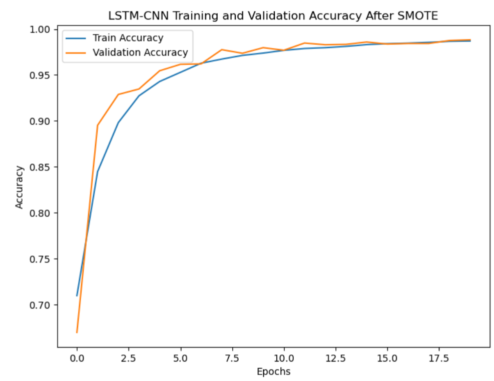

Figure 6: Training and Validation Accuracy for LSTM-CNN (3-Second Window, After SMOTE)

- Larger windows (10 seconds) led to a significant decline in LSTM-CNN performance, likely due to overfitting and the challenge of modeling extended temporal dependencies.
- Random Forest demonstrated robust performance with smaller windows (91% test accuracy) but lacked the capacity to model temporal dependencies effectively.
- Applying SMOTE enhanced model performance for minority classes, improving recall and F1-scores, particularly for underrepresented activities such as "Downstairs" and "Standing."

These findings emphasize the importance of selecting appropriate time-window sizes and leveraging temporal modeling techniques like LSTM-CNN for HAR systems. Future work could explore adaptive windowing techniques and lightweight architectures to optimize performance for real-time and edge-device applications.

A. Limitations and Future Directions

While this study provides valuable insights, it has certain limitations:

- The dataset used is limited to six activities, and additional activities might reveal different patterns.
- Larger window sizes showed decreased performance for LSTM-CNN, highlighting the need for adaptive windowing techniques to optimize both responsiveness and accuracy.
- Computational performance on resource-constrained edge devices was not evaluated.

Future work could focus on:

- Investigating dynamic window sizes based on activity types to improve both short-term and long-term recognition.
- Exploring lightweight models optimized for edge devices to enable real-time recognition in practical applications.
- Extending the analysis to multimodal datasets incorporating additional sensors like gyroscopes and magnetometers.

REFERENCES

[1] Ya Min, Yin Yin Htay, and Khin Khin Oo. Comparing the performance of machine learning algorithms for human activities recognition using wisdm dataset. International Journal of Computer (IJC), 38(1):61–72, 2020.

[2] Kun Xia, Jianguang Huang, and Hanyu Wang. Lstm-cnn architecture for human activity recognition. IEEE Access, 8:56855–56866, 2020.

[3] Pardeep Seelwal, Chetana Srinivas, VS Rekha, et al. Human activity recognition using wisdm datasets. Journal of Online Engineering Education, 14(1s):88–94, 2023.

[4] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors, 16(1):115, 2016.

[5] Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang. Deep learning with long short-term memory for time series prediction. IEEE Communications Magazine, 57(6):114–119, 2019.

[6] Tonny I Okedi and Adrian C Fisher. Time series analysis and long short-term memory (lstm) network prediction of bpv current density. Energy Environmental Science, 14(4):2408–2418, 2021.

[7] Mohammadreza Heydarian and Thomas E Doyle. rwisdm: Repaired wisdm, a public dataset for human activity recognition. arXiv preprint arXiv:2305.10222, 2023.

[8] Preeti Agarwal and Mansaf Alam. A lightweight deep learning model for human activity recognition on edge devices. Procedia Computer Science, 167:2364–2373, 2020.