

An Open-Source, Open Data Approach to Activity Classification from Triaxial Accelerometry in an Ambulatory Setting

Sepideh Nikookar¹, Edward Tian², Harrison Hoffman²,
Matthew Parks², J. Lucas McKay^{1,3}, Yashar Kiarashi¹,
Tommy T. Thomas^{1,4}, Alex Hall⁵, David W. Wright⁵
and Gari D. Clifford^{1,6}

¹Department of Biomedical Informatics, Emory University, Atlanta, GA, USA

²LifeBell AI LLC, Atlanta, GA, USA

³Department of Neurology, Emory University, Atlanta, GA, USA

⁴Department of Neurosurgery, Emory University, Atlanta, GA, USA

⁵Department of Emergency Medicine, Emory University, Atlanta, GA, USA

⁶Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

E-mail: gari@gatech.edu

Abstract.

The accelerometer has become an almost ubiquitous device, providing enormous opportunities in healthcare monitoring beyond step counting or other average energy estimates in 15-60 second epochs.

Objective: to develop an open data set with associated open-source code for processing 50 Hz tri-axial accelerometry-based to classify patient activity levels and natural types of movement.

Approach: Data were collected from 23 healthy subjects (16 males and seven females) aged between 23 and 62 years using an ambulatory device, which included a triaxial accelerometer and synchronous lead II equivalent ECG for an average of 26 minutes each. Participants followed a standardized activity routine involving five distinct activities: lying, sitting, standing, walking, and jogging. Two classifiers were constructed: a signal processing technique to distinguish between high and low activity levels and a convolutional neural network (CNN)-based approach to classify each of the five activities.

Main results: The binary (high/low) activity classifier exhibited an F1 score of 0.79. The multi-class CNN-based classifier provided an F1 score of 0.83. The code for this analysis has been made available under an open-source license together with the data on which the classifiers were trained and tested.

Significance: The classification of behavioral activity, as demonstrated in this study, offers valuable context for interpreting traditional health metrics and can be integrated into clinical decision-making tools for patient monitoring, predictive analytics, and personalized health interventions.

Keywords: 3D Accelerometry; Activity Recognition; Wearable Sensors; Signal Processing; Machine Learning.

1. Introduction

Continuous monitoring of physiological parameters plays a vital role in modern healthcare, enabling earlier detection of clinical deterioration and supporting personalized treatment strategies. Traditional biomedical monitoring approaches often relied on sporadic measurements collected during clinical visits or hospitalizations. However, advances in sensing technologies have increasingly enabled continuous, real-time data collection [Graña Possamai et al. \(2020\)](#); [Goldsack et al. \(2020\)](#); [Teixeira et al. \(2021\)](#), potentially transforming both clinical and home-based patient care.

These technological innovations span two major domains: on-body and off-body sensors. On-body wearable devices, such as smartwatches [Jat and Grønli \(2022\)](#) and in-ear EEG systems [Joyner et al. \(2024\)](#); [Waters et al. \(2024\)](#), capture physiological signals directly from the body. Off-body sensors, including thermal cameras [Manullang et al. \(2021\)](#) and radar systems [Fraccaro et al. \(2020\)](#), offer non-contact methods for monitoring patients from a distance. Alternative non-contact modalities, such as vision-based [Sathyaranayana et al. \(2018\)](#), audio-based [Carrón et al. \(2021\)](#), and radar/LiDAR-based systems [Li et al. \(2013\)](#); [Dong et al. \(2024\)](#); [Rinchi et al. \(2023\)](#), have been explored to detect activity patterns and vital signs without requiring physical contact.

Despite these advancements, non-contact monitoring faces critical limitations in real-world healthcare environments. Vision-based systems are sensitive to occlusions and lighting variability, while radar- and LiDAR-based methods, though promising, often suffer from environmental clutter and privacy concerns. These challenges can limit the reliability of non-contact approaches, particularly in dynamic and crowded clinical settings.

As a result, wearable sensors remain the most practical and scalable solution for continuous patient monitoring in many environments. Accelerometers, in particular, have garnered significant attention for their ability to provide real-time insights into patient activity and movement patterns. Accelerometers are widely used in remote patient monitoring, and to some extent in emergency department waiting rooms [Sweeney et al. \(2010\)](#); [Fung et al. \(2015\)](#); [Curtis et al. \(2008\)](#); [Nino et al. \(2020\)](#), and home-based care settings [Albahri et al. \(2019\)](#); [Wang et al. \(2017\)](#). By capturing continuous data on patient motion, accelerometers offer valuable context for interpreting physiological parameters such as heart rate and oxygen saturation [Ivașcu and Negru \(2021\)](#). For example, knowing whether a patient is walking, standing, or lying down can help clinicians accurately attribute changes in heart rate, reducing the risk of false alarms or misinterpretations.

However, even wearable sensors are not immune to challenges. Motion artifacts caused by natural patient movements can introduce noise into telemetry data, complicating real-time assessments [Freeman et al. \(2006\)](#); [Pawar et al. \(2007\)](#). Movements such as walking, shifting posture, or adjusting clothing can interfere with the accurate monitoring of heart rate, respiratory rate, and oxygen saturation. Although studies have validated the reliability of wearable devices compared to traditional clinical

monitors, demonstrating comparable heart rate and respiratory rate measurements in intensive care unit settings Stevens et al. (2024), real-world wearable data often suffer from noise, missingness, and imbalanced class distributions Taffoni et al. (2018). These data quality issues can degrade model performance, particularly in ambulatory health monitoring applications where environmental variability is high.

Human activity recognition (HAR) plays a crucial role in mitigating these challenges by providing context for interpreting physiological data. However, existing HAR methods face several limitations. Many models struggle with small labeled datasets, difficulty capturing temporal dependencies, computational constraints for real-time deployment, and the impact of noise on recognition accuracy Arshad et al. (2022). While architectures such as LSTM-based models have shown promise Mekruksavanich and Jitpattanakul (2021); Reyes-Ortiz et al. (2013), their effectiveness often relies on high-quality, synchronized time-series data, which may not be available in real-world ambulatory settings. Similarly, multi-sensor fusion strategies—such as combining accelerometry with ECG data—have improved classification performance in specific contexts Ren et al. (2024), but require complex hardware configurations that may not be feasible in resource-limited environments.

To address these limitations, this study focuses on developing robust, scalable activity recognition methods based solely on 3D accelerometry data. CNN-based approaches have demonstrated strong potential for human activity recognition across a variety of domains, capturing temporal and spatial patterns in sequential sensor data Shi et al. (2023); Ismail et al. (2023); Bao and Intille (2004); Yuan et al. (2022); Zhu et al. (2018). These strategies aim to balance classification accuracy, computational efficiency, and practical feasibility for real-world deployment in ambulatory health monitoring systems.

While prior studies, such as Ren et al. (2024), demonstrated strong performance using accelerometer-based models, their datasets Ren (2023) were collected in controlled laboratory environments with balanced data distributions and longer per-participant recordings. In contrast, this work targets the challenges of noisy, imbalanced data collected from shorter, more variable ambulatory recordings, which more closely reflect the data characteristics expected in real-world health monitoring systems. Rather than aiming to outperform models trained on high-fidelity, controlled datasets, the focus of this study is on developing robust methods that generalize under the imperfect and unpredictable conditions that commonly arise in practical healthcare deployments.

Similarly, Luckhurst et al. (2024) developed accelerometer-based cut-points for binary activity classification using a Mean Amplitude Deviation (MAD) metric. In contrast, this study applies a rolling 3D magnitude method that can be implemented in near real-time, producing classification windows with a fixed delay of $(N - 1)/2$ samples. For the 5 s window at 50 Hz ($N = 250$ samples), this corresponds to a delay of approximately 2.5 s, which could be considered negligible for real-time applications. While MAD captures variability around the mean acceleration, the 3D magnitude reflects overall movement intensity and is easier to interpret in streaming applications.

The threshold used in this study was determined empirically by analyzing distributions from early pilot data collected before the main study. This value was then spot-checked against the study dataset to ensure it generalized well to the broader population.

The objective of this study was to develop open data and evaluate open-source accelerometry-based methods for accurately classifying activity levels and specific types of movement that patients may exhibit in routine ambulatory situations. By automatically creating activity-type labels, we may improve the contextual interpretation of physiological signals, such methods can enhance clinical decision-making, reduce false alarms, and support more personalized health interventions in diverse healthcare environments.

To this end, two complementary approaches were implemented: a real-time, a signal processing-based method to distinguish between high and low activity levels through cut-point-based classification, and a CNN-based deep learning model for multi-class activity recognition. While the cut-point-based method is computationally efficient and suitable for real-time deployment, the deep learning model enables more granular and clinically meaningful behavioral assessments.

To the best of our knowledge, this is the first work to release both open source code and open access data for triaxial accelerometry-based human activity classification, thus enabling full reproducibility of our work.

2. Methods

2.1. Device Details

A practical and reusable wearable ECG monitor cardiac patch (the Vivalink VV330 Continuous ECG Platform and VivaLNK Adhesive Patch) was used for data collection. This patch serves as a versatile tool for gathering physiological data. It records single lead ECG data at a rate of 128 Hz, providing detailed insights into heart activity. Additionally, it captures triaxial accelerometry data at a rate of 50 Hz, which is the sampling rate provided by the device manufacturer, to offer valuable information about body movements and activity levels. Generally, human movement is confined to 20 Hz or less [Antonsson and Mann \(1985\)](#). A Nyquist frequency of 25 Hz is therefore sufficient to capture all relevant movement.

Furthermore, the patch automatically calculates heart rate measurements at a frequency of 1 Hz based on the recorded ECG signals. This feature enables real-time monitoring of heart rate variations, enhancing the depth of information collected during the study.

Although the device collects both ECG and accelerometry signals, only the accelerometry data were utilized for activity classification in this study. The ECG signals were analyzed separately to assess physiological parameters, such as heart rate deviations, to quantify the physiology of the population we used, but they were not fused with motion data for activity recognition. While both data streams were available from

the same hardware unit, the computational and modeling simplicity of our approach stems from using only accelerometry for real-time activity classification. This separation ensures that the proposed activity recognition methods remain scalable to much cheaper devices that collect only accelerometry.

2.2. Data Collection Protocol

Participants were recruited through flyers distributed in the office of Prophecy Games, an online computer game company affiliated with one of the authors, and in neighboring offices within the same building. This company provided only the office space for flyer distribution and had no role in the study design, data collection, analysis, or interpretation; there is no financial or scientific conflict of interest related to this arrangement. These flyers were approved by the Institutional Review Board (IRB) and invited individuals to participate in the study. A total of 23 participants (16 males and seven females), aged between 23 and 62 years, were enrolled. Relevant physiology (age, weight, and height) were recorded. However, demographic variables such as race and education level were not collected, which limits the ability to assess model generalizability across different populations.

The mean age of participants was 32.44 ± 9.84 years for males and 36.00 ± 12.01 years for females. All participants provided informed consent prior to participation, and the study received ethical approval from WCG IRB (Pr. No.: 20231639). Each participant engaged in a structured series of activities to replicate common scenarios encountered in non-sedentary hospital environments, such as those in emergency departments. These activities included a range of movements and postures typical of such environments, ensuring a representative dataset for analysis.

Prior to data collection, subjects underwent a standard skin cleaning procedure with an alcohol swab to optimize contact between the ECG patch electrodes and the skin. Each subject was fitted with a single patch, securely applied to approximate the Lead II position on their left chest using a specially designed adhesive patch. This standardized application method ensured consistent electrode positioning across all subjects, which is crucial for reliable data collection. Since each participant's data collection occurred independently, any malfunction or error with an individual device would only affect that participant's data, while the data from other participants would remain unaffected. This study design ensured that the overall dataset remained robust despite potential issues with individual devices. To further enhance data quality, participants with excessive chest hair were shaved to ensure optimal contact between the patch and the skin; this was essential since excessive hair could disrupt data collection.

Before commencing the study, thorough checks were performed to confirm the smooth transmission of data from the patch to the recording system and temporal continuity of each data packet. This step was crucial to ensure the accuracy and consistency of the data, establishing a reliable basis for the subsequent analysis. Data were streamed to a smartphone via Bluetooth and offloaded to AWS for secure storage

and analysis.

Subjects were guided through various activities in an office environment that mirrored the dynamic nature of patient behavior in emergency department waiting areas. These activities included lying down, sitting, standing, walking, and jogging. While subjects sat in office chairs and laid on a couch during the sedentary tasks, walking and jogging activities were specifically conducted in a larger conference room, where subjects walked back and forth or in circles to further simulate real-world conditions. Walking in circles was included as a controlled way to simulate movement in a confined space, which can be representative of certain real-world conditions, particularly in environments where space is limited, such as hospital waiting areas. This activity was intended to capture the physical exertion and movement patterns of individuals in dynamic, constrained spaces. While walking in circles may not universally represent all real-world scenarios, it provides a practical method to induce movement that mimics some aspects of real-world activity. Instructions were provided for participants to perform fidgeting behaviors, such as checking their mobile phones, bouncing their legs, retying their shoes, or adjusting their clothes or hair, to capture the subtleties of real-world scenarios.

To simulate potential disruptions, subjects were asked to tap on the patch or scratch the skin around it. This deliberate interference was designed to test the robustness of the monitoring system under challenging conditions, providing insights into its reliability and resilience in practical settings.

On average, each subject contributed 26.33 ± 21.36 minutes of continuous ECG and accelerometry data throughout the activity protocol. The minimum duration of data collected from an individual participant was 4.22 minutes, while the maximum duration was 67.98 minutes. The large standard deviation reflects the variation in participation, as some subjects were not comfortable completing all activities outlined in the protocol, resulting in shorter data collection duration. This data was further complemented by synchronized video recordings, which served as ground truth references for accurate labeling and validation. Videos of the subjects' activities were recorded on an iPad, and the iPad and patch were aligned with the world clock before every session to ensure that the video and accelerometry data were as time-aligned as possible. The combination of video and physiological data facilitated a comprehensive analysis of motion-related artifacts, enabling a thorough investigation into how motion affects biomedical data interpretation.

2.3. Labeling

The activity data collected was carefully labeled by hand, with close reference to the ground truth videos, to ensure accuracy. To label the data, an open-source tool, Label Studio, was used, which allowed for collaboration among team members and provided transparency in the labeling process. While only two individuals were responsible for the actual labeling, the platform remained accessible to the entire research team, allowing

for oversight and review by qualified personnel, including a registered nurse. The data was initially grouped into different activity states: lying, sitting, standing, walking, and jogging. Each of these states was considered separately, without any overlap between them.

The distribution of these activity states across the dataset is shown in Figure 1. This pie chart visually represents the proportions of time each subject spent in each activity during the data collection process, providing insights into the relative frequency of each activity.

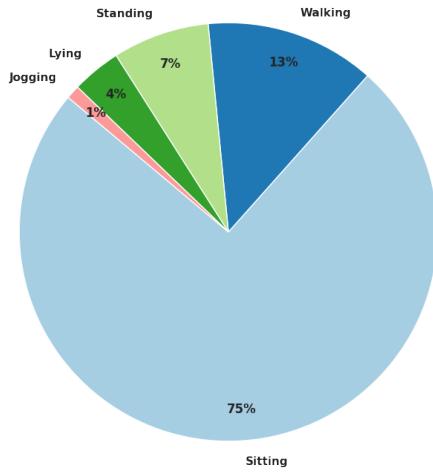


Figure 1. Activity label distribution across the dataset.

In parallel, the data were categorized based on the subjects' posture. This included whether they were leaning forward, leaning backward, leaning to the left, or leaning to the right. If a specific posture wasn't identified, it was assumed the subjects were upright.

In addition to categorizing activity and posture, labels were assigned to capture general activities during the data collection process. The label assigned to time t was whatever activity the subject performed in the window from $t - 5$ to t seconds. These labels encompassed transitions between different postures, instances where the device might have been interfered with, and other actions like fidgeting or adjusting one's position. The 5-second window was selected to balance capturing sufficient temporal context for each activity while maintaining precision in detecting activity transitions. Larger window sizes could lead to diminishing returns by smoothing out subtle transitions, making classification less accurate.

While the vast majority of windows only contained a single activity, as intended by the protocol, there were windows where the subject transitioned between activities. In such cases, the window was labeled according to the activity that occurred most recently within the window. If a clear transition between two activities was observed,

the window was labeled as a transition. Transition-labeled windows were excluded from subsequent model training and evaluation to ensure that the models were trained on well-defined, stable activity states.

Importantly, the activity labels (lying, sitting, standing, walking, and jogging) were mutually exclusive , meaning that a window could only be labeled with one of these activity states at a time. While participants were instructed to perform natural fidgeting behaviors, these micro-movements were not assigned independent labels in the final dataset. Instead, windows containing fidgeting were labeled based on the primary activity the subject was engaged in at the time. This decision was made to ensure consistency in activity classification, as fidgeting often overlapped with multiple activity states and did not represent a distinct category within the model framework.

2.4. Preprocessing

To prepare the data for analysis, triaxial accelerometry signals collected from each subject were utilized. The preprocessing steps were crucial to ensure the data was clean and reliable for subsequent analysis.

First, all three axes of the accelerometry signals were preprocessed by a Butterworth bandpass filter with cutoff frequencies set at 0.05 Hz and 2 Hz. The low-frequency cutoff of 0.05 Hz removed sensor drift, while the high-frequency cutoff of 2 Hz filtered out irrelevant signals, such as high-frequency vibrations from devices like phones or nearby machinery. Studies have shown that most human motion-related signals, including walking, running, sitting, and standing, predominantly fall below 20 Hz, as reported in frequency domain analyses of human activity recognition [Antonsson and Mann \(1985\)](#). To further validate this choice, a sensitivity analysis was conducted by varying the high-frequency cutoff to 2, 5, 10, 15, and 20 Hz. The results of this analysis are presented in Section 3.2.3. This filtering approach preserved essential physiological or motion-related signals, such as walking, running, and even rapid leg bouncing. The Butterworth filter was chosen for its maximally flat frequency response in the passband. It does not introduce sharp transitions or excessive distortion, preserving the data integrity while filtering out noise.

To ensure consistency in the dataset and exclude non-protocol-related activity, only the last 20 minutes of accelerometry data for each subject were included in the analysis. For participants with less than 20 minutes of data, all their data were retained for analysis. This approach aimed to exclude any potential noise, such as preparation or calibration periods at the beginning of the recording, and focus the analysis on the structured activity protocol. Importantly, no specific activities were intentionally removed; instead, this step was taken to enhance the quality and relevance of the dataset.

2.5. Classification Approaches

2.5.1. Cut-Point Based Classification. As a first approach to characterizing motion, the aim was to classify activity levels into a binary outcome: active or inactive.

This distinction is a critical step in interpreting biomedical data such as ECG and heart rate. Activity levels influence the interpretation of heart rate variability, as the expected ranges for heart rate and ECG patterns differ based on whether a subject is at rest or in motion. Moreover, it is critical to understand if changes in HRV are endogenous (related to a change in physiology) or exogenous (due to a change in physical activity). Therefore, distinguishing between active and inactive states provides valuable metadata to inform clinical assessments.

To accomplish this, the vector magnitude (also known as the Euclidean norm) of the accelerometry data was computed. This calculation combined the filtered signals from the x, y, and z axes into a single value at each time point using the formula $\sqrt{x_{\text{filtered}}^2 + y_{\text{filtered}}^2 + z_{\text{filtered}}^2}$. This provided a measure of the overall movement intensity, taking into account the contributions of all three axes of the accelerometry signal.

To classify activity levels, a rolling threshold approach was applied to the vector magnitude values. A rolling threshold involves evaluating the accelerometer signal magnitude over a moving time window. At each time point, the rolling median of the magnitude values is calculated within a specified 5-second window. This method helps smooth out short-term fluctuations and highlights trends over time, facilitating a more accurate classification of activity states. The 5-second window was selected based on its ability to capture the dynamic nature of human movements while maintaining a balance between temporal resolution and signal stability. Previous studies have shown that mid-sized windows size (from 5 to 7 seconds long) allow for capturing sufficient movement characteristics of an activity while also enabling the quick detection of transitions between activities, striking a balance between detailed analysis and real-time responsiveness [Twomey et al. \(2018\)](#); [Janidarmian et al. \(2017\)](#). A rolling median was specifically chosen because it is more robust to outliers in the accelerometry signal, such as sudden spikes caused by device artifacts or environmental vibrations, ensuring a reliable signal for subsequent analysis and classification.

2.5.2. Multi-Class Classification. For the second approach, a machine learning methodology was employed to classify multiple distinct activities based on accelerometry data. A 1-dimensional Convolutional Neural Network (CNN) was chosen due to its ability to effectively capture temporal patterns and hierarchical features within sequential data, such as accelerometry time series. CNNs have been widely used in time-series classification tasks, as they are capable of learning spatial and temporal correlations in sequential data, making them well-suited for detecting activity-related patterns [Hammerla et al. \(2016\)](#); [Chen et al. \(2021\)](#). This approach aimed to recognize and differentiate between five activities: lying, sitting, standing, walking, and jogging.

As detailed in Section 2.4, all accelerometry signals were first preprocessed using a Butterworth bandpass filter with cutoff frequencies of 0.05 Hz and 2 Hz to remove drift and high-frequency noise. To validate the robustness of this choice, a sensitivity analysis was performed by systematically varying the high-frequency cutoff to 5, 10, 15, and 20 Hz, as described in Section 3.2.3. This analysis was conducted to ensure the

filtering strategy preserved the most relevant human activity signals while minimizing irrelevant noise.

The CNN architecture used in this study is illustrated in Figure 2, detailing the convolutional, pooling, and dense layers. The network was trained using data labeled with the five activity classes. During training, the model learned to associate specific spatiotemporal patterns in the accelerometry signals with each class.

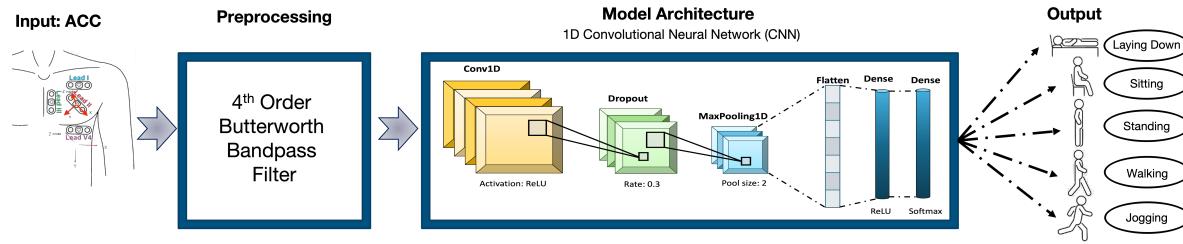


Figure 2. Architecture of the deep learning model for the activity classification.

Model performance was evaluated using Leave-One-Out Cross-Validation (LOOCV). In this setup, data from one participant was held out for testing while the model was trained on the remaining participants. This process was repeated iteratively for all participants, ensuring that no individual's data was seen during training on the fold where they were tested. This approach provides a rigorous evaluation of generalization performance in small datasets.

To quantify model performance, accuracy, precision, recall, and F1-score were computed using weighted averaging. This approach assigns importance to each class based on its prevalence, thereby accounting for class imbalance and ensuring that metrics are not dominated by more frequent activity types. Alternative methods, such as macro and micro averaging, were considered but found less appropriate given the highly imbalanced nature of the dataset (e.g., overrepresentation of sitting and underrepresentation of jogging). Weighted metrics therefore offer a more realistic estimate of model utility in clinical or real-world deployments where minority classes are equally important.

To evaluate how model performance is affected by different parameters, input window sizes (1s, 2s, 4s, 5s, 8s, 12s) and sampling rates (5 Hz, 10 Hz, 25 Hz, 50 Hz) were systematically varied. While a 5-second window and 50 Hz sampling rate were justified as the default based on prior literature and empirical optimization, exploring a broader range of parameter settings provides important insights into model robustness, especially under real-time or resource-constrained scenarios. These analyses support the development of adaptable systems that can maintain accuracy under varying computational or sensor limitations.

To further assess whether observed differences in performance were attributable to experimental design choices or participant-level variability, a linear mixed-effects model was applied. The outcome metric (e.g., F1-score) was modeled as a function of window

size and sampling rate, which were treated as fixed effects due to their systematic manipulation across experiments. Subject ID was modeled as a random effect to account for inter-individual differences in activity performance or sensor placement. This modeling approach allowed us to disentangle the influence of these design parameters from subject-specific noise and better evaluate their generalizability. Performance scores from all window-sampling rate combinations were included as input to the model. Full results are reported in Section 3.2.

2.6. Resting Vitals

This analysis builds upon the cut-point based classification method described in Section 2.5.1 by evaluating the physiological relevance of the rolling 3D magnitude signal. Specifically, the analysis examined whether greater movement intensity, as captured by the 5-second rolling magnitude, is associated with deviations from each subject's resting heart rate. This relationship supports the use of motion-derived features for contextualizing cardiovascular responses in ambulatory settings.

Each subject's resting heart rate was estimated by computing the median heart rate during periods labeled as 'lying' or 'sitting' prior to the first recorded instance of walking. Heart rate deviation was then calculated as the difference between each observed heart rate value and the individual's baseline. To quantify movement intensity, the rolling 3D magnitude measure derived from accelerometer data was used. Pearson correlation analysis was performed to evaluate the relationship between rolling magnitude and heart rate deviation, and a scatter plot was generated to visualize the trend. This approach provides insight into the potential of using physical activity features as proxies for physiological stress or exertion.

3. Results

3.1. Cut-Point Based Classification Performance

This section reports the performance of the cut-point based classification method in distinguishing active and inactive states using accelerometry data.

A threshold of 0.07 was set empirically by analyzing the distributions of early data collected prior to the main study. These preliminary observations showed a clear separation in rolling 3D magnitude values between inactive states (lying, sitting, standing) and active states (walking, jogging). To ensure generalizability, the threshold was spot-checked against a subset of the main study data and found to remain effective in separating activity states. The selected threshold was then applied to the full dataset to classify activity levels as either inactive or active.

Subsequently, 80% of the subjects from the main dataset (with no participant overlap between training and testing sets) were used to assess the performance of this threshold.

Model performance was assessed using a held-out validation set comprising 20% of participants. The model achieved an accuracy of 0.87, a precision of 0.73, and a recall of 0.90. This highlights the model's ability to accurately distinguish activity levels under typical ambulatory conditions.

3.1.1. Distribution Analysis. The distribution of 3D magnitude values was compared between inactive and active states in the training set.

Figure 3 presents the distribution of rolling magnitude values, computed as the rolling median of the 3D acceleration vector magnitude over a 5-second window. This representation allows for distinguishing inactive states (lying, sitting, standing) and active states (walking, jogging) using the proposed threshold of 0.07. Standing was categorized as an inactive state following prior literature [Kuster et al. \(2020\)](#), which suggests that standing does not necessarily involve substantial movement and is often grouped with other inactive states in accelerometry-based studies.

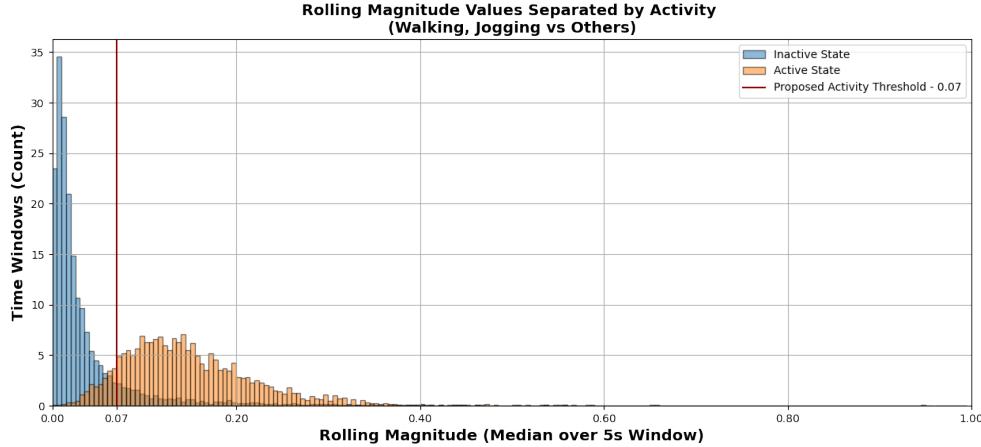


Figure 3. Histogram of Activity Count Values Separated by Activity.

3.1.2. Confusion Matrix. The confusion matrix, shown in Figure 4, provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. This analysis offers more profound insight into the model's performance, highlighting areas of strength and potential misclassifications.

3.1.3. Comparison with MAD-Based Cut-Point. To assess how existing methods generalize to the current dataset, the Mean Amplitude Deviation (MAD)-based cut-point proposed by [Luckhurst et al. \(2024\)](#) was applied. Their study identified an optimal MAD threshold of 47.73 mG (0.04773 g) for classifying sedentary and ambulatory activity using the same VivaLink ECG Patch.

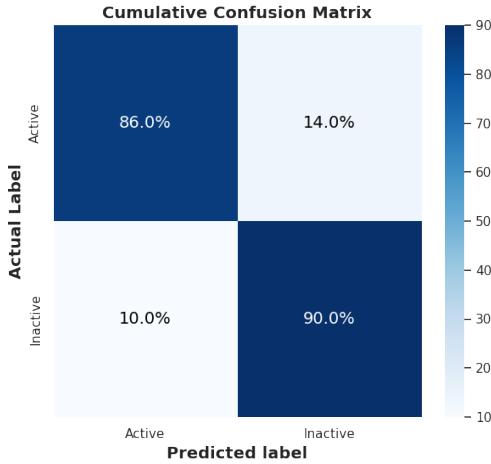


Figure 4. Confusion Matrix for Active vs. Inactive Labels.

When this threshold was applied to the current dataset using a 5-second rolling MAD window, the method achieved an accuracy of 0.71. However, it failed to identify any active windows, resulting in a precision and recall of 0.00. This suggests the MAD cut-point is overly conservative in real-world ambulatory data and misclassifies most active states as inactive.

In contrast, the 3D magnitude-based threshold of 0.07 achieved substantially higher performance across all metrics highlighting its robustness in noisier, more variable datasets. These results support the need for context-specific methods—including both the choice of activity metric and thresholding strategy—that are resilient to the variability and noise inherent in real-world ambulatory monitoring settings.

3.2. Multi-Class Classification Performance

This section presents the results of the CNN-based multi-class activity classification using filtered triaxial accelerometry data. The model was trained to classify five distinct activity types—lying, sitting, standing, walking, and jogging—using LOOCV for evaluation. Performance was assessed under various preprocessing configurations to determine the impact of window size, sampling rate, and filtering parameters.

3.2.1. Baseline Model Performance. The baseline model was evaluated using a 5-second input window and a 50 Hz sampling rate. As summarized in Table 1, accuracy, precision, recall, and F1-score were computed using weighted averaging. The model achieved a weighted accuracy of 77%, precision of 83%, recall of 77%, and F1-score of 79%. These metrics reflect the model’s ability to generalize across participants.

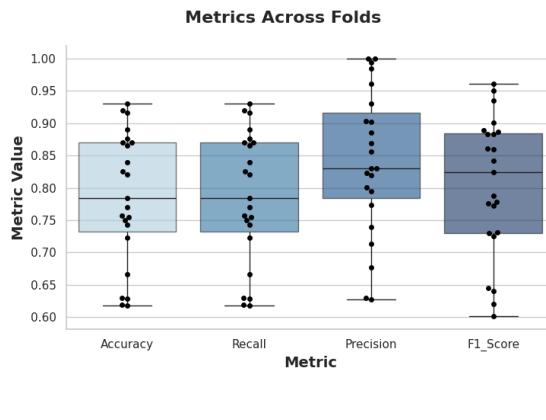
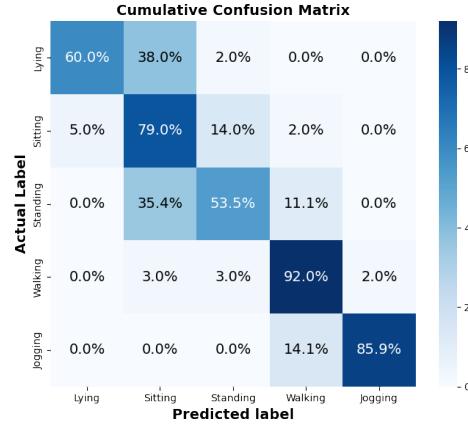
3.2.2. Distribution and Confusion Matrix. To assess consistency across participants, box plots of evaluation metrics are shown in Figure 5. The relatively

Metric	Weighted Score
Accuracy	77%
Precision	83%
Recall	77%
F1 Score	79%

Table 1. Weighted performance metrics of the CNN classifier.

narrow distributions confirm stable performance of the CNN across individuals.

A confusion matrix was also included to further illustrate the model's classification performance. This matrix, presented in Figure 6, highlights that the base model tends to misclassify somewhat similar activities. Notably, the model frequently confuses sitting and standing, as both postures exhibit similar vertical (z-axis) fluctuations, making them challenging to differentiate with a single accelerometer. In many instances, standing is misclassified as walking, particularly when the 5-second window includes minor walking movements or other motions that contaminate the standing data. However, the model does not confuse more distinct activities, such as walking and lying down, which indicates its ability to differentiate between activities with varying motion patterns.

**Figure 5.** Box Plots of Different Metrics Across Different Folds for 5s Window Size.**Figure 6.** Confusion Matrix for 5s Window Size.

3.2.3. Sensitivity of CNN Classification to Cutoff Frequencies. To examine how the CNN classification performance responds to variations in the filtering step described in Section 2.4, a sensitivity analysis was conducted by varying the high-frequency cutoff in the Butterworth filter (2, 5, 10, 15, and 20 Hz). As shown in Figure 7, classification performance declined as the cutoff frequency increased, suggesting that frequencies above 2 Hz introduce noise without improving discriminatory information. These results support the original filter design and demonstrate that lower cutoff frequencies preserve relevant motion signals for activity classification.

This experiment was conducted using the multi-class CNN model to ensure that conclusions about filter design directly inform model performance.

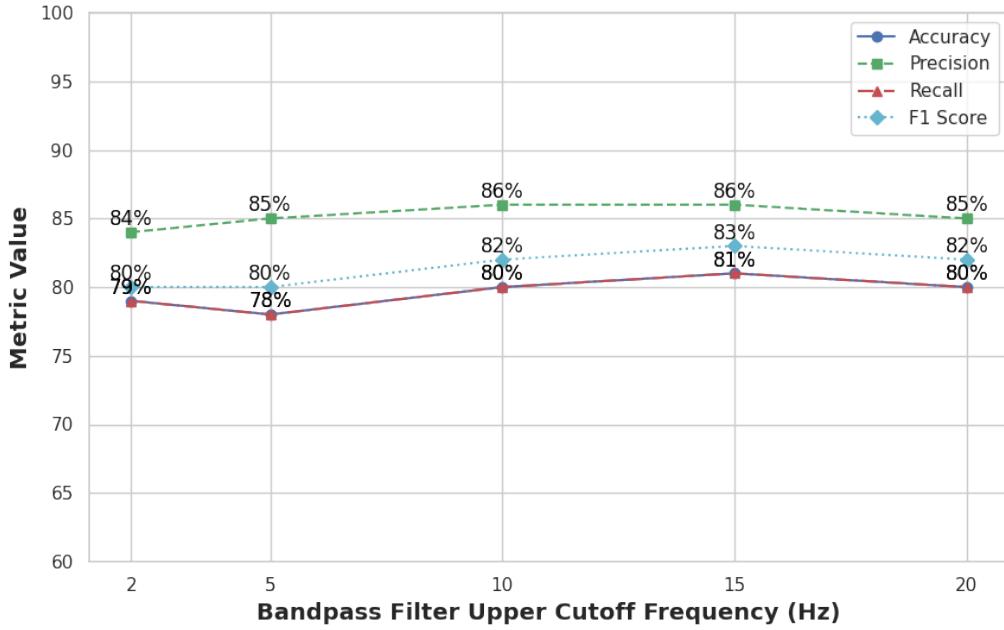


Figure 7. Classification performance across different high-frequency cutoff values in the Butterworth filter.

3.2.4. Impact of Varying Input Size and Sampling Rate. In this section, the effect of varying the window size and sampling rate on the model’s performance was investigated. The tested window sizes were 1, 2, 4, 5, 8, and 12 seconds, and the sampling rates were 5 Hz, 10 Hz, 25 Hz, and 50 Hz. The performance metrics—accuracy, recall, precision, and F1 score—generally decreased with shorter window sizes and lower sampling rates, likely due to the limited temporal context available in smaller windows. This reduction in information adversely affected the model’s ability to accurately classify activities.

To visualize these impacts, heatmaps were generated for each metric (Figure 8), showing how different combinations of window size and sampling rate influence model performance. These plots provide insights into the trade-offs between temporal resolution and classification accuracy.

3.2.5. Mixed-Effect Model Analysis. This section presents the results of a mixed-effect model analysis to assess whether the observed performance differences are influenced by subject-specific factors or driven primarily by experimental parameters. The model quantified the impact of window size and sampling rate on classification

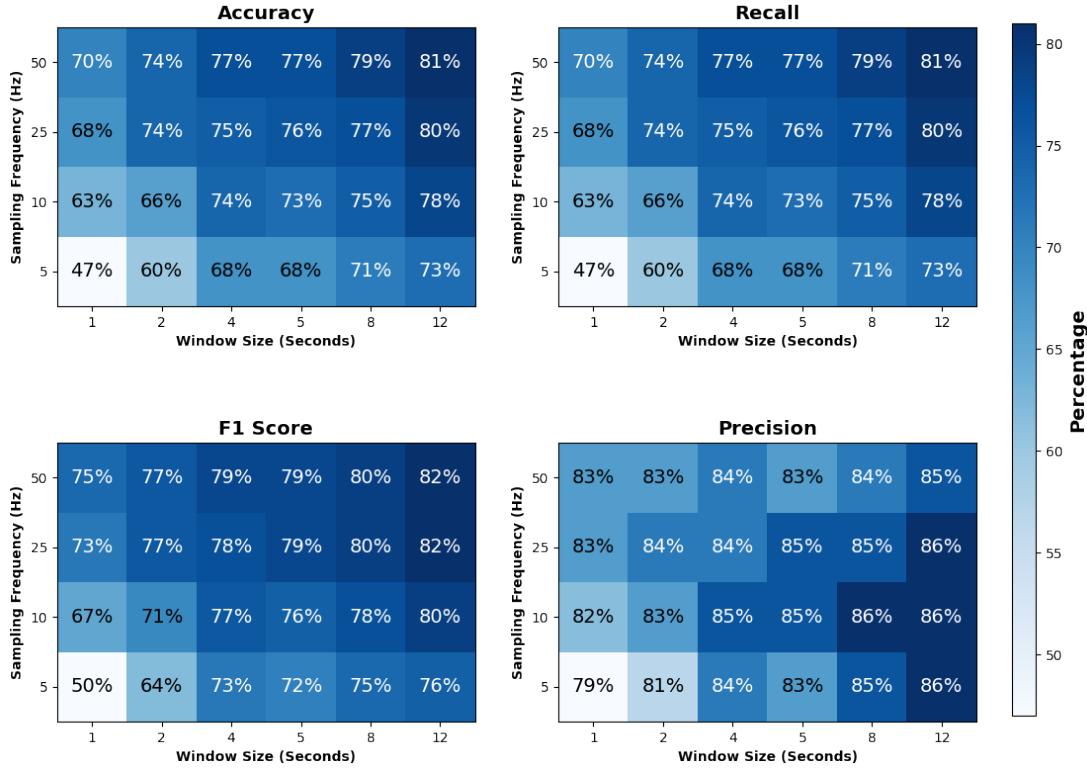


Figure 8. Heatmap of Different Metric vs. Window Size and Sampling Rate.

metrics while controlling for inter-subject variability. Table 2 summarizes the estimated coefficients and confidence intervals.

The results indicate that reducing the window size from 12 seconds to shorter durations leads to significant decreases in classification performance. For example, 1-second windows resulted in the largest negative effect ($\beta = -12.389$, $p < 0.001$). Similarly, decreasing the sampling rate from 50 Hz to 5 Hz or 10 Hz substantially reduced performance, with the most pronounced effect observed at 5 Hz ($\beta = -8.786$, $p < 0.001$).

In contrast, changing the evaluation metric revealed that precision tends to score higher than F1, while both accuracy and recall tend to score lower, relative to F1-score. The random effect variance for subject ID (70.490) confirms the presence of inter-individual variability, though the fixed effects dominate the performance trends. These results reinforce that window size and sampling rate are primary drivers of model performance, beyond subject-specific differences.

Predictor	Coef.	Std. Err.	z	$P > z $	[0.025, 0.975]
Window Size 1 s vs. 12 s	-12.389	0.594	-20.853	0.000	[-13.553, -11.224]
Window Size 2 s vs. 12 s	-7.264	0.594	-12.226	0.000	[-8.428, -6.099]
Window Size 4 s vs. 12 s	-3.516	0.594	-5.919	0.000	[-4.681, -2.352]
Window Size 5 s vs. 12 s	-3.120	0.594	-5.251	0.000	[-4.284, -1.955]
Window Size 8 s vs. 12 s	-1.736	0.594	-2.923	0.003	[-2.901, -0.572]
Sampling Rate 5 Hz vs. 50 Hz	-8.786	0.485	-18.113	0.000	[-9.737, -7.835]
Sampling Rate 10 Hz vs. 50 Hz	-3.361	0.485	-6.928	0.000	[-4.314, -2.409]
Sampling Rate 25 Hz vs. 50 Hz	-0.783	0.485	-1.613	0.107	[-1.733, 0.168]
Accuracy vs. F1 Score	-3.284	0.485	-6.771	0.000	[-4.235, -2.334]
Precision vs. F1 Score	8.333	0.485	18.333	0.000	[7.942, 8.724]
Recall vs. F1 Score	-3.284	0.485	-6.771	0.000	[-4.235, -2.334]
Group Var	70.490	2.675			

Table 2. Mixed Linear Model Regression Results

3.3. Resting Vitals Analysis

In this section, the observed relationship between movement intensity—quantified by the rolling 3D magnitude—and heart rate deviation from resting baseline—is reported. A Pearson correlation analysis revealed a statistically significant positive association ($r = 0.29$, $p < 0.0001$), indicating that higher levels of physical activity correspond to greater deviations from a subject’s resting heart rate.

Figure 9 illustrates this relationship, where each point represents a 5-second window during an active period.

4. Discussion

The findings highlight the potential of 3D accelerometry for activity classification in ambulatory monitoring, with direct applications in healthcare diagnostics and patient monitoring. The ability to distinguish between different activity states, as demonstrated by the machine learning and signal processing approaches proposed in this study, is crucial for improving movement analysis in rehabilitation, chronic disease management, and elder care. For example, the results show that heart rate deviations correlate with physical activity, reinforcing the need to integrate motion tracking with physiological monitoring for more accurate health assessments. This finding aligns with previous studies that emphasize the value of contextualizing physiological data using motion sensors, such as in heart rate and oxygen saturation interpretation Ivașcu and Negru (2021); Freeman et al. (2006). This capability is particularly relevant for conditions like cardiovascular disease and congestive heart failure, where physical activity plays a critical role in disease management Greiwe and Nyenhuis (2020).

Although real-time classification was not implemented or tested in this study, all analyses were conducted retrospectively. Nevertheless, the methods demonstrated here

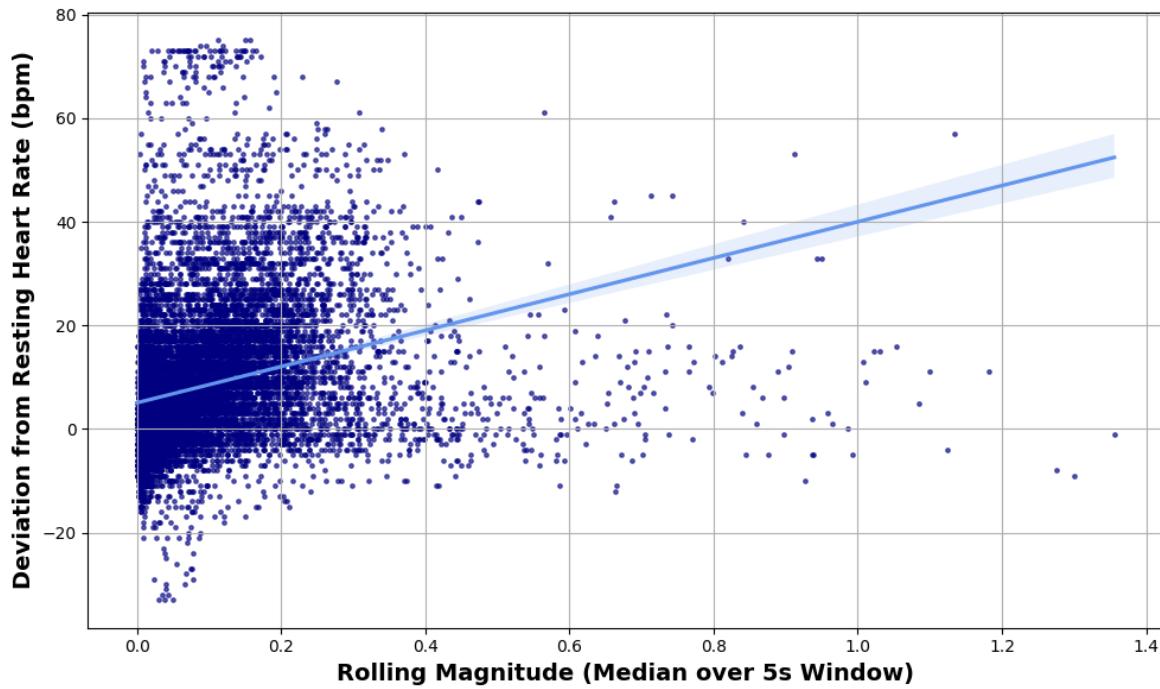


Figure 9. Relationship between Activity Count and Heart Rate Deviation.

provide a foundation for future development of real-time systems, as they use a 5 s sliding window with low per-update computational cost for the rolling-magnitude method ($\sim O(N)$ operations) and millisecond-scale inference time for the CNN framework. These findings suggest that accelerometry-based models could potentially be incorporated into continuous monitoring applications in clinical or home settings. Such integration is especially pertinent given the growing reliance on wearable technologies in remote and ambulatory care environments Jat and Grønli (2022); Albahri et al. (2019); Stevens et al. (2024). It is important to note that the model was not trained on transitional periods and is therefore rather unstable when presented with transitional data. The rapid fluctuations are typical of any hysteresis at a margin of transition. It is therefore recommended that periods of rapid fluctuations in the model output are labelled as ‘transition’ or unknown’.

Building on these findings, the combined use of 3D accelerometry data with signal processing and machine learning techniques has the potential to enhance the ability to manage variability in ambulatory settings caused by patient movement. The signal processing approach provides a computationally efficient method for distinguishing general activity levels, while the machine learning model, particularly the CNN, enables more detailed and precise activity classification, allowing for a more nuanced understanding of patient behavior. This supports prior work showing CNNs are well-suited for time-series accelerometry data in HAR applications Shi et al. (2023); Yuan et al. (2022).

A key finding of this study is that heart rate deviations correlate with physical activity, reinforcing the need to integrate motion tracking with physiological monitoring for more accurate health assessments. The resting vitals analysis underscores the importance of contextualizing physiological measurements within activity states. By establishing a subject's resting heart rate based on periods of inactivity (lying or sitting), deviations in heart rate due to movement were quantified. This approach ensures a more accurate interpretation of cardiovascular responses in ambulatory settings.

Without accounting for activity levels, heart rate monitoring may misinterpret physiological stress, leading to potential false alarms or misdiagnoses. For example, a retrospective analysis may show that an increase in heart rate could be misclassified as a stress response, when it may simply reflect a transition from rest to movement. The findings reinforce the need to integrate motion tracking with heart rate monitoring to enhance the clinical utility of wearable health devices Pawar et al. (2007).

The results demonstrate that combining 3D accelerometry data with signal processing and machine learning techniques enhances the ability to manage variability in ambulatory settings caused by patient movement. The signal processing approach provides a computationally efficient method for distinguishing general activity levels, while the CNN-based model enables detailed multi-class activity classification. Although the current study does not evaluate live or embedded classification, these methods are designed with computational efficiency in mind and could be adapted for future real-time deployments. These complementary methods allow for both future real-time monitoring and more nuanced behavioral assessments, which are critical for personalized health interventions.

While this study does not present a complete biomedical monitoring system, it demonstrates the feasibility of using accelerometry-based CNN models for accurate multi-class activity classification. Unlike device-specific pipelines that rely on idealized datasets or multi-sensor inputs, the studied approach was developed using noisy, imbalanced data collected under realistic conditions. This design choice directly addresses common limitations of HAR systems outlined in prior studies, including challenges with data quality, noise, and generalizability Taffoni et al. (2018); Arshad et al. (2022); Ren et al. (2024).

Despite these promising findings, practical challenges remain in balancing model complexity with computational efficiency. The choice of window size and sampling rate significantly influences model performance, with longer windows and higher sampling rates generally yielding higher classification accuracy. However, these improvements come at the cost of increased computational demands, which may limit feasibility for real-time or resource-constrained applications. In this study, the observed performance gains with larger windows were likely driven by long stretches of homogeneous activity, which are easier to classify. In operational settings, however, longer windows may also capture transitions between activities—such as standing and walking—leading to ambiguity and potential misclassification. Future efforts should aim to optimize this trade-off, ensuring both accuracy and efficiency in more complex, real-world conditions.

One potential solution to this issue is to assign the activity label based on the most recent portion of the window. This approach assumes that the last segment of the window reflects the most relevant context for prospective applications or downstream decision-making. Since transitions often involve mixed activity patterns, focusing on the final portion of the window may help reduce misclassification due to overlapping behaviors, while still preserving the performance benefits of longer input sequences.

Despite the strengths of this study, several limitations must be acknowledged. For example, the study focused on a single accelerometer placement, which may limit the ability to differentiate between specific postures (e.g., sitting vs. standing). Future research should investigate multi-sensor approaches, such as utilizing accelerometers on the torso and thigh, to further enhance posture and activity detection Wang et al. (2017). Expanding the dataset to encompass a broader range of populations and real-world scenarios will also improve generalizability and clinical applicability.

Another limitation involves the treatment of ambiguous activity periods. In this study, transition windows and interference-labeled segments were excluded from training and evaluation to ensure well-defined class boundaries. While this approach improved classification performance for stable activity states, it limits the model's applicability in handling ambiguous but common real-world movements. Future work should explore strategies for detecting or modeling such transitional states directly, which could enhance the completeness and robustness of real-world activity recognition Arshad et al. (2022).

Moreover, an important limitation of this study is the absence of collected demographic variables such as race and education level. While the study's primary aim was to develop and validate activity recognition methods, which are not expected to be primarily influenced by race (for example), the size of the cohort and the lack of demographic data prevent assessment of model generalizability across such factors. Future studies should prioritize collecting comprehensive demographic information to enable evaluation of potential biases and to improve the equity and applicability of activity classification models.

All data, code, trained models, and supporting documentation used in this research have been made publicly available Nikookar et al. (2025).

5. Conclusion

This study successfully evaluated two approaches for recognizing patient activity using 3D accelerometry sensor data to improve the utility of biomedical monitoring in ambulatory settings. The signal processing approach effectively distinguished between high and low activity levels, while the machine learning algorithm demonstrated strong performance in classifying five distinct activities: lying down, sitting, standing, walking, and jogging. By accurately identifying activity states, these methods have the potential to help contextualize physiological measurements, reduce motion-induced inaccuracies and support the interpretation of wearable sensor data in healthcare applications. Additionally, the states themselves may offer diagnostic insights, as factors like time

spent in a specific state or the frequency of transitions between states could be explored in future work for links to conditions such as over-sedation, delirium, or sleep-related illnesses. The proliferation of low cost and low-energy accelerometer devices, combined with advanced activity recognition techniques, holds promise to provide critical context for biomedical parameters, potentially enabling more reliable health monitoring in dynamic healthcare environments.

Acknowledgments

This study was partially funded by LifeBell AI LLC. Dr. Clifford holds equity in LifeBell. LifeBell has an interest in the type of software technology being developed and evaluated in the research described in this paper. The terms of this arrangement have been reviewed and approved by Emory University in accordance with its conflict of interest policies. Dr. Clifford is also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR002378 and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH award number R01EB030362. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, LifeBell AI LLC, or the authors' current and past employers and funding bodies. This study was approved by WCG IRB (Pr. No.: 20231639).

References

- Albahri, A. S., Albahri, O. S., Zaidan, A., Zaidan, B., Hashim, M., Alsalem, M., Mohsin, A. H., Mohammed, K., Alamoodi, A. H., Enaizan, O. et al. (2019). Based multiple heterogeneous wearable sensors: A smart real-time health monitoring structured for hospitals distributor, *IEEE Access* **7**: 37269–37323.
- Antonsson, E. K. and Mann, R. W. (1985). The frequency content of gait, *Journal of biomechanics* **18**(1): 39–47.
- Arshad, M. H., Bilal, M. and Gani, A. (2022). Human activity recognition: Review, taxonomy and open challenges, *Sensors* **22**(17): 6463.
- Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data, *International Conference on Pervasive Computing*, Springer, pp. 1–17.
- Carrón, J., Campos-Roca, Y., Madruga, M. and Pérez, C. J. (2021). A mobile-assisted voice condition analysis system for Parkinson’s disease: Assessment of usability conditions, *Biomedical Engineering Online* **20**: 1–24.
- Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z. and Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities, *ACM Computing Surveys (CSUR)* **54**(4): 1–40.
- Curtis, D. W., Pino, E. J., Bailey, J. M., Shih, E. I., Waterman, J., Vinterbo, S. A., Stair, T. O., Guttag, J. V., Greenes, R. A. and Ohno-Machado, L. (2008). SMART—an integrated wireless system for monitoring unattended patients, *Journal of the American Medical Informatics Association* **15**(1): 44–53.
- Dong, S., Wen, L., Ye, Y., Zhang, Z., Wang, Y., Liu, Z., Cao, Q., Xu, Y., Li, C. and Gu, C. (2024). A review on recent advancements of biomedical radar for clinical applications, *IEEE Open Journal of Engineering in Medicine and Biology* **5**: 707–724.
- Fraccaro, P., Evangelopoulos, X. and Edwards, B. (2020). Development and preliminary evaluation of a method for passive, privacy-aware home care monitoring based on 2d lidar data, *Artificial Intelligence in Medicine: 18th International Conference on Artificial Intelligence in Medicine, AIME 2020, Minneapolis, MN, USA, August 25–28, 2020, Proceedings 18*, Springer, pp. 160–169.
- Freeman, J. V., Dewey, F. E., Hadley, D. M., Myers, J. and Froelicher, V. F. (2006). Autonomic nervous system interaction with the cardiovascular system during exercise, *Progress in Cardiovascular Diseases* **48**(5): 342–362.
- Fung, E., Järvelin, M., Doshi, R., Shinbane, J., Carlson, S., Grazette, L., Chang, P., Sangha, R., Huikuri, H. and Peters, N. (2015). Electrocardiographic patch devices and contemporary wireless cardiac monitoring, *Frontiers in Physiology* **6**.
- Goldsack, J., Coravos, A., Bakker, J., Bent, B., Dowling, A., Fitzer-Attas, C., Godfrey, A., Godino, J., Gujar, N., Izmailova, E., Manta, C., Peterson, B., Vandendriessche, B., Wood, W., Wang, K. and Dunn, J. (2020). Verification, analytical validation, and clinical validation (v3): the foundation of determining fit-for-purpose for biometric monitoring technologies (biomets), *NPJ Digital Medicine* **3**.

- Graña Possamai, C., Ravaud, P., Ghosn, L. and Tran, V.-T. (2020). Use of wearable biometric monitoring devices to measure outcomes in randomized clinical trials: a methodological systematic review, *BMC Medicine* **18**: 1–11.
- Grewe, J. and Nyenhuis, S. (2020). Wearable technology and how this can be implemented into clinical practice, *Current Allergy and Asthma Reports* **20**.
- Hammerla, N. Y., Halloran, S. and Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables, *arXiv preprint arXiv:1604.08880*.
- Ismail, W. N., Alsalamah, H. A., Hassan, M. M. and Mohamed, E. (2023). AUTO-HAR: An adaptive human activity recognition framework using an automated cnn architecture design, *Heliyon* **9**(2).
- Ivașcu, T. and Negru, V. (2021). Activity-aware vital sign monitoring based on a multi-agent architecture, *Sensors* **21**(12): 4181.
- Janidarmian, M., Roshan Fekr, A., Radecka, K. and Zilic, Z. (2017). A comprehensive analysis on wearable acceleration sensors in human activity recognition, *Sensors* **17**(3): 529.
- Jat, A. S. and Grønli, T.-M. (2022). Smart watch for smart health monitoring: a literature review, *International Work-Conference on Bioinformatics and Biomedical Engineering*, Springer, pp. 256–268.
- Joyner, M., Hsu, S.-H., Martin, S., Dwyer, J., Chen, D. F., Sameni, R., Waters, S. H., Borodin, K., Clifford, G. D., Levey, A. I. et al. (2024). Using a standalone ear-eeg device for focal-onset seizure detection, *Bioelectronic Medicine* **10**(1): 4.
- Kuster, R. P., Grootenhuis, W. J., Blom, V., Baumgartner, D., Hagströmer, M. and Ekblom, Ö. (2020). Is sitting always inactive and standing always active? a simultaneous free-living activpal and actigraph analysis, *International journal of environmental research and public health* **17**(23): 8864.
- Li, C., Lubecke, V. M., Boric-Lubecke, O. and Lin, J. (2013). A review on recent advances in doppler radar sensors for noncontact healthcare monitoring, *IEEE Transactions on Microwave Theory and Techniques* **61**(5): 2046–2060.
- Luckhurst, J., Hughes, C. and Shelley, B. (2024). Classifying physical activity levels using mean amplitude deviation in adults using a chest worn accelerometer: validation of the vivalink ECG patch, *BMC Sports Science, Medicine and Rehabilitation* **16**(1): 212.
- Manullang, M. C. T., Lin, Y.-H., Lai, S.-J. and Chou, N.-K. (2021). Implementation of thermal camera for non-contact physiological measurement: A systematic review, *Sensors* **21**(23): 7777.
- Mekruksavanich, S. and Jitpattanakul, A. (2021). LSTM networks using smartphone data for sensor-based human activity recognition in smart homes, *Sensors* **21**(5): 1636.
- Nikookar, S., Tian, E., Hoffman, H. and Clifford, G. D. (2025). Code and data for

- activity classification from triaxial accelerometry. Accessed: 2025-02-11.
URL: https://github.com/cliffordlab/Accel_Activity_DB/
- Nino, V., Claudio, D., Schiel, C. and Bellows, B. (2020). Coupling wearable devices and decision theory in the united states emergency department triage process: A narrative review, *International Journal of Environmental Research and Public Health* **17**(24): 9561.
- Pawar, T., Anantakrishnan, N., Chaudhuri, S. and Duttagupta, S. P. (2007). Impact analysis of body movement in ambulatory ECG, *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 5453–5456.
- Ren, Y. (2023). Human activity recognition acc electrocardiogram, figshare. Dataset.
URL: <https://doi.org/10.6084/m9.figshare.24132510.v3>
- Ren, Y., Liu, M., Yang, Y., Mao, L. and Chen, K. (2024). Clinical human activity recognition based on a wearable patch of combined tri-axial acc and ecg sensors, *Digital Health* **10**: 20552076231223804.
- Reyes-Ortiz, J., Anguita, D., Ghio, A., Oneto, L. and Parra, X. (2013). Human activity recognition using smartphones, UCI Machine Learning Repository. Dataset.
URL: <https://doi.org/10.24432/C54S4K>
- Rinchi, O., Ghazzai, H., Alsharoa, A. and Massoud, Y. (2023). Lidar technology for human activity recognition: Outlooks and challenges, *IEEE Internet of Things Magazine* **6**(2): 143–150.
- Sathyaranayana, S., Satzoda, R. K., Sathyaranayana, S. and Thambipillai, S. (2018). Vision-based patient monitoring: a comprehensive review of algorithms and technologies, *Journal of Ambient Intelligence and Humanized Computing* **9**: 225–251.
- Shi, L.-F., Liu, Z.-Y., Zhou, K.-J., Shi, Y. and Jing, X. (2023). Novel deep learning network for gait recognition using multimodal inertial sensors, *Sensors* **23**(2): 849.
- Stevens, G., Larmuseau, M., Damme, A. V., Vanoverschelde, H., Heerman, J. and Verdonck, P. (2024). Feasibility study of the use of a wearable vital sign patch in an intensive care unit setting, *Journal of Clinical Monitoring and Computing* pp. 1–12.
- Sweeney, K., Leamy, D., Ward, T. and McLoone, S. (2010). Intelligent artifact classification for ambulatory physiological signals.
- Taffoni, F., Rivera, D., La Camera, A., Nicolò, A., Velasco, J. R. and Massaroni, C. (2018). A wearable system for real-time continuous monitoring of physical activity, *Journal of Healthcare Engineering* **2018**(1): 1878354.
- Teixeira, E., Fonseca, H., Diniz-Sousa, F., Veras, L., Bopp, G., Oliveira, J., Pinto, D., Alves, A., Barbosa, A., Mendes, R. and Marques-Aleixo, I. (2021). Wearable devices for physical activity and healthcare monitoring in elderly people: a critical review, *Geriatrics* **6**: 38.
- Twomey, N., Diethe, T., Fafoutis, X., Elsts, A., McConville, R., Flach, P. and Craddock,

- I. (2018). A comprehensive study of activity recognition using accelerometers, *Informatics*, Vol. 5, MDPI, p. 27.
- Wang, Z., Yang, Z. and Dong, T. (2017). A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time, *Sensors* **17**(2): 341.
- Waters, S., Berent, J., Saini, P. and Clifford, G. (2024). Domain adaptation using large scale databases for sleep staging from a novel in-ear sensor.
- Yuan, L., Andrews, J., Mu, H., Vakil, A., Ewing, R., Blasch, E. and Li, J. (2022). Interpretable passive multi-modal sensor fusion for human identification and activity recognition, *Sensors* **22**(15): 5787.
- Zhu, Q., Chen, Z. and Soh, Y. C. (2018). A novel semisupervised deep learning method for human activity recognition, *IEEE Transactions on Industrial Informatics* **15**(7): 3821–3830.