Activity Classification from Triaxial Accelerometry in an Ambulatory Setting

Sepideh Nikookar[1], Edward Tian[2], Harrison Hoffman[2], Matthew Parks[2], J. Lucas McKay[1,3], Yashar Kiarashi[1], Tommy T. Thomas[1,4], Alex Hall[5], David W. Wright[5] and Gari D. Clifford[1,6]

[1]Department of Biomedical Informatics, Emory University, Atlanta, GA, USA
[2]LifeBell AI LLC, Atlanta, GA, USA
[3]Department of Neurology, Emory University, Atlanta, GA, USA
[4]Department of Neurosurgery, Emory University, Atlanta, GA, USA
[5]Department of Emergency Medicine, Emory University, Atlanta, GA, USA
[6]Department 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. This study explored the use of processing 50Hz triaxial accelerometry sensor data to identify patient activity.

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 for classifying each of the five activities. The objective of this study was to evaluate and compare the performance of these classifiers in accurately identifying and distinguishing between the activity levels and specific activities. 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 have been made available under an open-source license.

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. Background

Biomedical monitoring plays a vital role in modern healthcare, driven by advancements in both on-body Bonato (2003); Ajerla et al. (2019) and off-body Poncette et al. (2020); Kiarashi et al. (2024) sensing technologies. These innovations have transformed clinical monitoring by enabling continuous data collection rather than relying on sporadic measurements Graña Possamai et al. (2020); Goldsack et al. (2020); Teixeira et al. (2021). On-body sensors, such as smartwatches Jat and Grønli (2022) and in-ear EEGs Joyner et al. (2024); Waters et al. (2024), capture physiological data directly, while off-body sensors, like thermal cameras Manullang et al. (2021) and radar systems Fraccaro et al. (2020), provide non-intrusive monitoring in clinical and home settings.

Wearable sensors, particularly accelerometers, are widely used in ambulatory settings such as remote patient monitoring, real-time tracking 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 Albahri et al. (2019); Wang et al. (2017). The ability to continuously monitor patient movement enhances decision-making in triage settings Ivașcu and Negru (2021). However, motion artifacts—caused by natural patient movement—introduce variability into biomedical data, complicating real-time assessments and potentially leading to false alarms or misinterpretations Freeman et al. (2006); Pawar et al. (2007). Movements such as walking, standing, or adjusting posture can interfere with telemetry signals, affecting the accuracy of physiological measurements such as heart rate and oxygen saturation.

While traditional biomedical monitoring methods rely primarily on contact-based sensors, alternative non-contact modalities such as vision-based Sathyanarayana et al. (2018), audio-based Carrón et al. (2021), and radar/LiDAR-based Li et al. (2013); Dong et al. (2024); Rinchi et al. (2023) systems are being explored for patient monitoring. Vision-based systems use RGB and infrared cameras to track movement patterns, while radar and LiDAR technologies enable real-time monitoring without physical contact. Despite these advancements, wearable accelerometers remain the most practical solution in dynamic environments where non-contact methods may be unreliable due to occlusion, lighting conditions, or privacy concerns.

Recent research has aimed to improve the reliability of wearable monitoring systems. Studies have validated wearable devices against traditional clinical monitors, demonstrating comparable accuracy for heart rate and respiratory rate measurements in ICU settings Stevens et al. (2024). The use of Wearable Activity Monitors (WAMs), such as accelerometers and gyroscopes, provides valuable context by distinguishing motion-induced noise from genuine physiological changes Maher et al. (2021). However, a persistent challenge in wearable monitoring is the presence of missing or imbalanced data, which can degrade model performance in real-world applications Taffoni et al. (2018).

Several studies have explored sensor fusion to enhance activity classification. Ren et al. (2024) Ren et al. (2024) demonstrated that combining accelerometry with ECG data improves classification accuracy. However, their model relies on multiple sensor modalities, which may not always be available in resource-limited settings. Additionally, their dataset Ren (2023) does not account for imbalanced data distributions or missing values, common challenges in wearable monitoring. Similarly, LSTM-based models have shown promising results for activity recognition in smart home environments Mekruksavanich and Jitpattanakul (2021); Reyes-Ortiz et al. (2013), but their performance depends on high-quality time-series data from well-calibrated sensors. These limitations highlight the need for more robust and adaptable approaches that can effectively handle data noise, missing values, and real-world variability—a challenge that this study aims to address.

2. Introduction

Accurately monitoring patient activity and physiological parameters in dynamic environments remains a major challenge in biomedical sensing. Traditional monitoring systems assume that patients remain stationary, but real-world healthcare settings—including emergency departments, hospital wards, and home-based care—involve continuous patient movement. These movements introduce motion-induced artifacts, which can obscure physiological signals, leading to inaccurate assessments, false alarms, and compromised clinical decision-making.

Human activity recognition (HAR) plays a crucial role in mitigating these challenges, but several limitations persist. Existing HAR methods often struggle with: (1) limited labeled datasets, (2) difficulty capturing temporal dependencies, (3) computational constraints, and (4) the impact of noise on recognition accuracy Arshad et al. (2022). Additionally, multi-sensor approaches, while effective, require extensive infrastructure, making them impractical for ambulatory and resource-limited settings.

To address these limitations, this study aims to develop and evaluate robust activity recognition methods using 3D accelerometry data. The objective is to determine an effective approach for classifying activity levels that can be integrated into biomedical monitoring systems while balancing computational efficiency and classification accuracy. Specifically, two complementary approaches are evaluated: 1) A signal processing method to distinguish between high and low activity levels using low-pass filtering and thresholding to reduce motion artifacts; 2) A machine learning-based classifier that detects multiple distinct activities using advanced pattern recognition techniques Shi et al. (2023); Ismail et al. (2023); Bao and Intille (2004); Yuan et al. (2022); Zhu et al. (2018).

Unlike prior studies that rely on multi-sensor fusion, this work exclusively utilizes accelerometer data, making it more practical for low-resource settings. Additionally, rolling median filtering is applied to suppress noise while minimizing computational overhead, and a 5-second rolling window is used to capture activity transitions more effectively. To further improve classification accuracy, synchronized video recordings were used during data labeling, reducing annotation errors that could impact model performance.

This study compares the effectiveness of these two approaches in distinguishing activity levels and improving the contextual interpretation of biomedical data. While machine learning models excel in adaptability and generalization, signal processing techniques provide greater explainability and computational efficiency. Evaluating both methods under real-world conditions provides insights into the trade-offs between accuracy, interpretability, and computational feasibility in ambulatory health monitoring.

This paper is structured as follows: Section 3 describes the methodology, including the data collection process, signal processing techniques, and machine learning approaches used for activity classification. Section 4 presents the results, evaluating the performance of both methods in distinguishing activity levels. Section 5 discusses the findings, addressing potential challenges and their implications for real-world biomedical monitoring. Finally, Section 6 provides the conclusion, summarizing key takeaways and suggesting directions for future research.

3. Methods

3.1. 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 frequecy 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.

3.2. Data Collection Protocol

Participants were recruited through flyers distributed in the office of a sister company and neighboring offices within the same building. 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 7 females), aged between 23 and 62 years, were enrolled. Relevant physiology (age, weight, and height) were recorded, but demographic information such as race and education level were not recorded since the cohort was not large enough to draw any conclusions about these variables. 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.

3.3. 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.

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

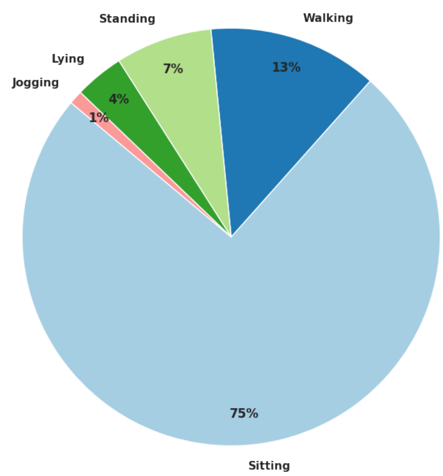

Figure 1. Activity label distribution across the dataset.

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. 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.

3.4. Prepocessing

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 4.2.1. 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.

3.5. Classification Approaches

3.5.1. Binary 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 $\sqrt{x^2_{\text{filtered}} + y^2_{\text{filtered}} + z^2_{\text{filtered}}}$. 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.

Using 80% of the data from subjects (with no overlap of participants between the training and testing sets), the rolling median values of the vector magnitude data between inactive states (lying, sitting, standing) and active states (walking, jogging) was compared. Analyzing the distribution of rolling magnitude values in the training set revealed a natural separation between active and inactive states, leading to the selection of 0.07 as an effective threshold for distinguishing between these categories. This threshold was then used to classify subjects' activity levels as either inactive or active.

This approach differs from prior studies, such as Luckhurst et al. (2024), which determined a Mean Amplitude Deviation (MAD)-based cut-point for classifying sedentary and ambulatory activity using the same VivaLink ECG Patch. Instead of MAD, this method utilizes rolling 3D magnitude, specifically developed for real-time ambulatory classification, making it more suitable for integration into biomedical monitoring systems. Additionally, while previous cut-points were optimized for general physical activity classification, the threshold in this study was designed to contextualize movement patterns in relation to physiological monitoring.

The binary classification model was evaluated using the remaining 20% of subjects' data as a validation set by applying the predefined 0.07 threshold to the smoothed vector magnitude values. Model performance was measured using accuracy, precision, recall, and the F1-score to ensure a balanced evaluation of classification effectiveness. Accuracy provided an overall measure of correct classifications, while precision and recall assessed the model's performance in identifying both active and inactive states. The F1-score, which combines precision and recall, was used to balance these two metrics, especially in cases where one class might be more prevalent than the other. This method provided a reliable way to classify activity levels in real-world ambulatory settings, helping to enhance the interpretation of biomedical data and account for different heart rate ranges based on whether the subject was active or inactive.

3.5.2. 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 focused on using the 1D CNN to recognize and differentiate between various postures and movements.

Like the binary classification approach, the accelerometry data first underwent preprocessing with a Butterworth bandpass filter. As detailed in Section 3.4, the filter used cutoff frequencies of 0.05 Hz and 2 Hz to remove low-frequency drift and high-frequency noise, respectively, while preserving human motion-related signals. The filtered accelerometry data was then used to train a 1-dimensional convolutional neural network (CNN) specifically designed to classify five different activities: lying, sitting, standing, walking, and jogging. Figure 2 illustrates the architecture of the CNN used in this study, detailing the specific layers and configurations employed to achieve the classification tasks.

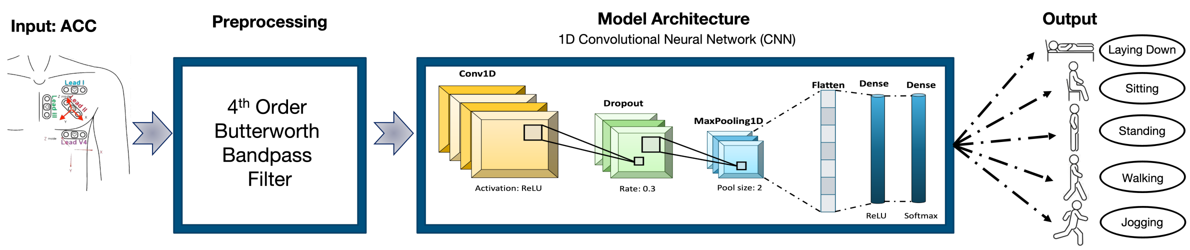

Figure 2. Architecture of the deep learning model for the activity classification.

The accelerometry data was labeled with the five activities mentioned above, and this labeled data was used to train the CNN. The model learned to associate specific patterns in the accelerometry signals with each activity, with the training process involving adjustment of the network parameters to minimize classification errors.

For evaluation, Leave-One-Out Cross-Validation (LOOCV) was used to assess the model's performance. In this approach, data from one participant was left out as the test set, while the model was trained on the data from all other participants. This process was repeated for each participant, ensuring no overlap between training and testing data. This approach allowed CNN to demonstrate its accuracy in distinguishing between five distinct activities and its ability to generalize to new, unseen data. The multi-class classification approach highlighted the effectiveness of using machine learning and CNNs for precise activity recognition based on accelerometry data, as evidenced by the results discussed in Section 4.2.

The development and evaluation of this 1-dimensional CNN model aimed to classify multiple human activities using accelerometry data. While a complete biomedical monitoring system has not yet been developed, this approach serves as a foundational step in demonstrating how machine learning models can be utilized for activity recognition. The results provide insights into the potential integration of machine learning into future biomedical monitoring systems for more accurate and reliable activity recognition.

3.6. Resting Vitals

To calculate each subject's resting heart rate, the median heart rate was analyzed while subjects were lying down or sitting. This approach provided a reliable estimate of baseline heart rate, free from the influence of higher activity levels such as walking or jogging. The impact of physical activity on cardiovascular response was assessed by analyzing deviations from each subject's resting heart rate in relation to movement intensity. Movement intensity was quantified using a rolling 3D magnitude measure, which reflects overall motion derived from accelerometer data. The purpose of this analysis was to assess whether variations in physical activity levels correspond to changes in heart rate, thereby evaluating the sensitivity of motion-based metrics in tracking cardiovascular responses.

For each subject, resting heart rate was determined using median heart rate values from data points labeled as 'lying' or 'sitting' before the first recorded instance of walking. Heart rate deviation was then calculated as the difference between the measured heart rate and the individual's resting heart rate. The relationship between movement and heart rate deviation was quantified by performing a Pearson correlation analysis between the rolling 3D magnitude and heart rate deviation. Pearson correlation was chosen because it measures the strength and direction of the linear relationship between two continuous variables. A scatter plot was generated to visualize the trend, providing insights into the potential use of motion-derived features as proxies for physiological stress.

4. Results

4.1. Binary Classification Performance

This section focuses on evaluating the binary classification model's ability to distinguish between active and inactive states using accelerometry data. The model's accuracy, precision, and recall were assessed based on a threshold derived from the 3D magnitude of the accelerometry signal. Additionally, distribution comparisons and a confusion matrix were used to provide a more comprehensive understanding of the classification outcomes.

Using the rolling 3D magnitude threshold of 0.07, the signal processing model's performance was assessed. In the validation set, the model achieved an accuracy of 0.81, a precision of 0.57, and a recall of 0.90. Transition states and interference introduce ambiguity in labeling, as these periods often exhibit mixed characteristics, making it difficult to confidently assign them to a single category. To ensure a clearer assessment of the model's ability to classify well-defined active and inactive states, these edge-case labels were excluded during performance evaluation. By removing these ambiguous cases, the metrics improved, with accuracy rising to 0.87, precision to 0.73, and recall remaining at 0.90. Such transition periods exist in real-world settings, and future work could explore strategies for handling these cases separately, rather than excluding them outright.

4.1.1. **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.

4.1.2. **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.

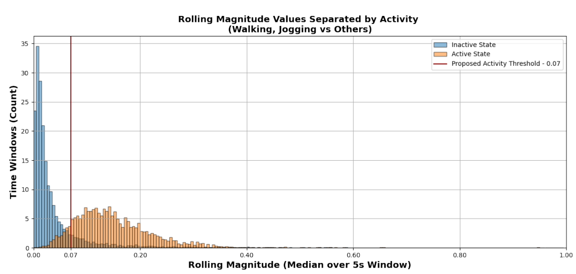

Figure 3. Histogram of Activity Count Values Separated by Activity.

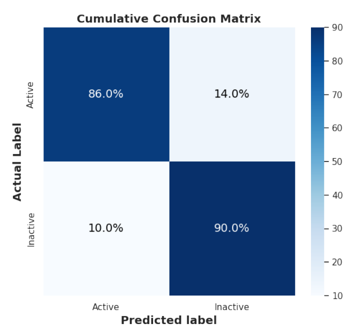

Figure 4. Confusion Matrix for Active vs. Inactive Labels.

4.2. Multi-Class Classification Performance

This section presents the performance of the multi-class classification model. First, the performance is discussed with a 5-second input window and the original 50 Hz sampling rate. Subsequently, the impact of varying these parameters on accuracy, recall, precision, and F1 score is examined by systematically adjusting the input window size (e.g., 1s, 2s, 4s, 5s, 8s and 12s) and the sampling rate (e.g., 5 Hz, 10 Hz, and 25 Hz). The classification model was trained and tested under each condition, and the resulting performance metrics were compared to assess how different preprocessing choices influence classification accuracy. This analysis provides insights into optimizing the model under various conditions.. The section begins with baseline performance metrics, followed by assessing parameter variations to provide insights for optimizing the model under various conditions.

The model's performance was first evaluated using a fixed window size of 5 seconds and a sampling rate of 50 Hz. Accuracy, precision, recall, and F1 score were calculated using weighted, macro, and micro-averaging methods. Table 1 summarizes these results.

Metrics / Type	**Weighted**	**Macro**	**Micro**
Accuracy	77%	77%	77%
Precision	83%	52%	77%
Recall	77%	81%	77%
F1-Score	79%	49%	77%

Table 1. Performance Comparison of Averaging Methods

The performance was assessed using three different averaging methods—weighted, macro, and micro—to provide a comprehensive comparison of the model's evaluation metrics. As shown in Table 1, all three averaging methods—weighted, macro, and micro—produced the same accuracy. However, the weighted metrics demonstrated higher precision and F1 scores than macro and micro averaging methods. While the macro averaging method achieved the highest recall, it was associated with lower precision and F1 scores, indicating a potential for false positives. Given the imbalanced nature of the dataset (see Figure 1), the weighted metrics were deemed the better option. This method ensures that the minority classes are appropriately weighted, providing a more balanced evaluation across all classes and mitigating the impact of class imbalance. Therefore, for the remainder of the experiments, weighted metrics were used, as they offer a better balance between key performance metrics and are more suitable for clinical applications, where handling class imbalance is critical for reliable and meaningful predictions.

Box plots for accuracy, recall, precision, and F1 score were generated to gain deeper insights into each metric's behavior. Figure 5 displays these box plots, with the x-axis representing the metrics and the y-axis representing the values. This visualization helps understand each metric's distribution and variability across different folds.

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.

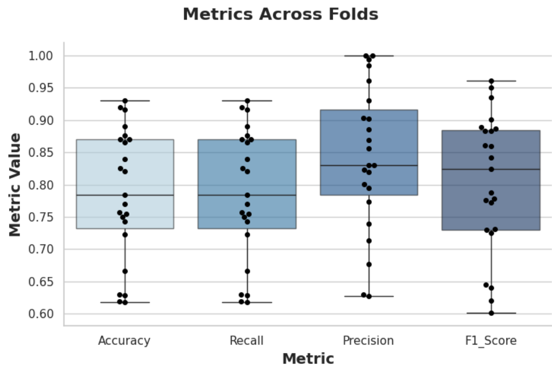

Figure 5. Box Plots of Different Metrics Across Different Folds for 5s Window Size.

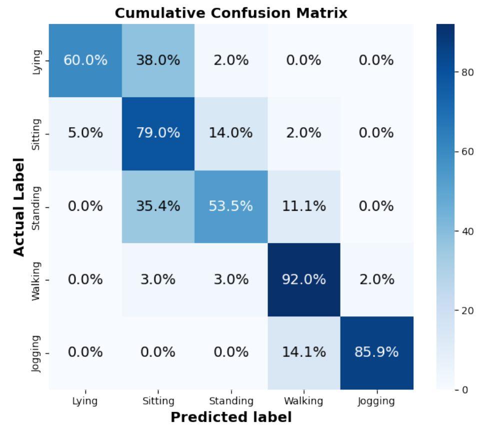

Figure 6. Confusion Matrix for 5s Window Size.

4.2.1. Sensitivity Analysis of Cutoff Frequencies In the preprocessing pipeline, a Butterworth bandpass filter with a high-frequency cutoff of 2 Hz was applied to remove irrelevant high-frequency noise while retaining motion-related signals. To evaluate the impact of this choice, a sensitivity analysis was conducted by varying the high-frequency cutoff to 2, 5, 10, 15, and 20 Hz, while assessing multi-class classification performance.

For each cutoff setting, the classification model was trained and tested, with accuracy, F1-score, precision, and recall compared. Figure 7 presents the performance across different cutoff values. The results indicate that increasing the high-frequency cutoff beyond 2 Hz did not significantly improve classification accuracy. Instead, higher cutoffs introduced additional high-frequency components, which did not contribute meaningfully to distinguishing activity classes.

Based on these findings, the 2 Hz cutoff was selected as it effectively removed high-frequency noise while maintaining robust classification performance.

4.2.2. 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 window sizes tested were 1 second, 2 seconds, 4 seconds, 5 seconds, 8 seconds, and 12 seconds, while the sampling rates tested were 5 Hz, 10 Hz, 25 Hz, and 50 Hz. The performance metrics—accuracy, recall, precision, and F1 score—tended to decrease with shorter windoe sizes and lower sampling rates due to the limited amount of information available within smaller windows. This leads to a loss of crucial temporal data, which diminishes the model's ability to classify activities accurately.

Although larger window sizes generally led to improved performance, this was not always the case, especially in real-world scenarios. The choice of window sizes and sampling rates was based on empirical experimentation and prior work, with considerations for real-time processing, computational efficiency, and the nature of the activities being classified. For instance, the 12-second window was selected as a

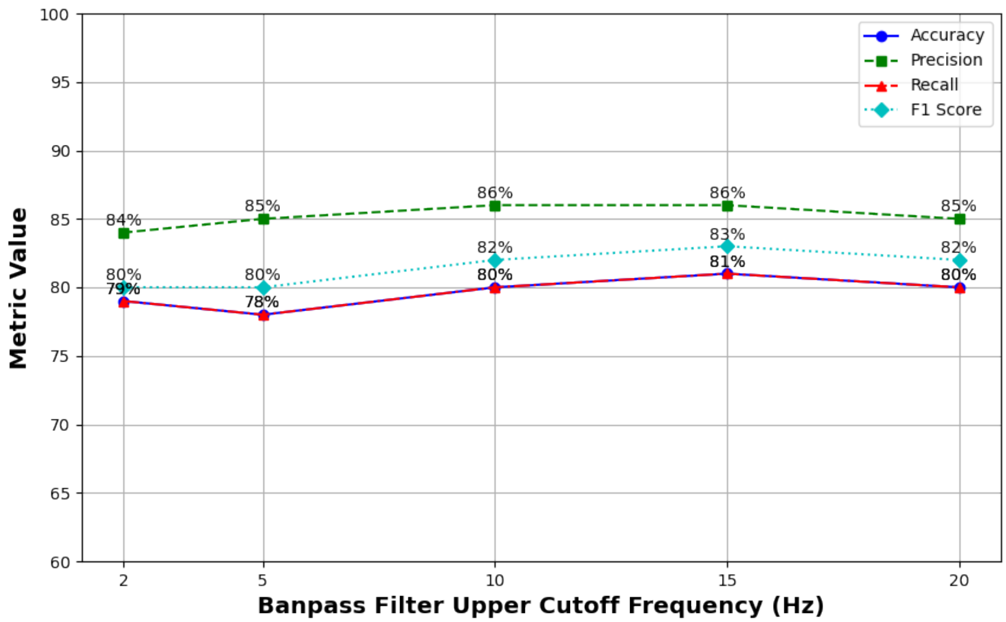

Figure 7. Classification performance across different high-frequency cutoff values in the Butterworth filter.

balance between capturing sufficient activity context and avoiding excessive computation time. The increase in performance observed in the dataset is mainly due to long segments of homogeneous activity, which are easier for the model to classify. In practice, however, larger windows often capture transitions between multiple activities, resulting in ambiguity and misclassification. For example, a window may contain both standing and walking, making it difficult to assign a single activity label. This is particularly problematic for activities like sitting and standing, which share similar postures but are distinct in other contexts, such as heart rate or movement patterns.

To visualize these impacts, heatmap plots were generated for each metric in Figure 8, illustrating how changes in the input size and sampling rate influenced the model's performance. These heatmaps provide valuable insights into the trade-offs involved, highlighting that while increasing window size can improve performance metrics, the potential for capturing heterogeneous activities must be considered.

One possible solution for handling larger windows is to assign the activity label based on the activity that occurred in the last portion of the window. This decision is based on the assumption that the most recent activity in the window is more relevant for real-time applications and downstream decision-making. Additionally, transitional periods often contain mixed activity patterns, which can introduce ambiguity and negatively impact model performance. By focusing on the final segment of the window, this approach aims to reduce misclassification errors caused by overlapping activities

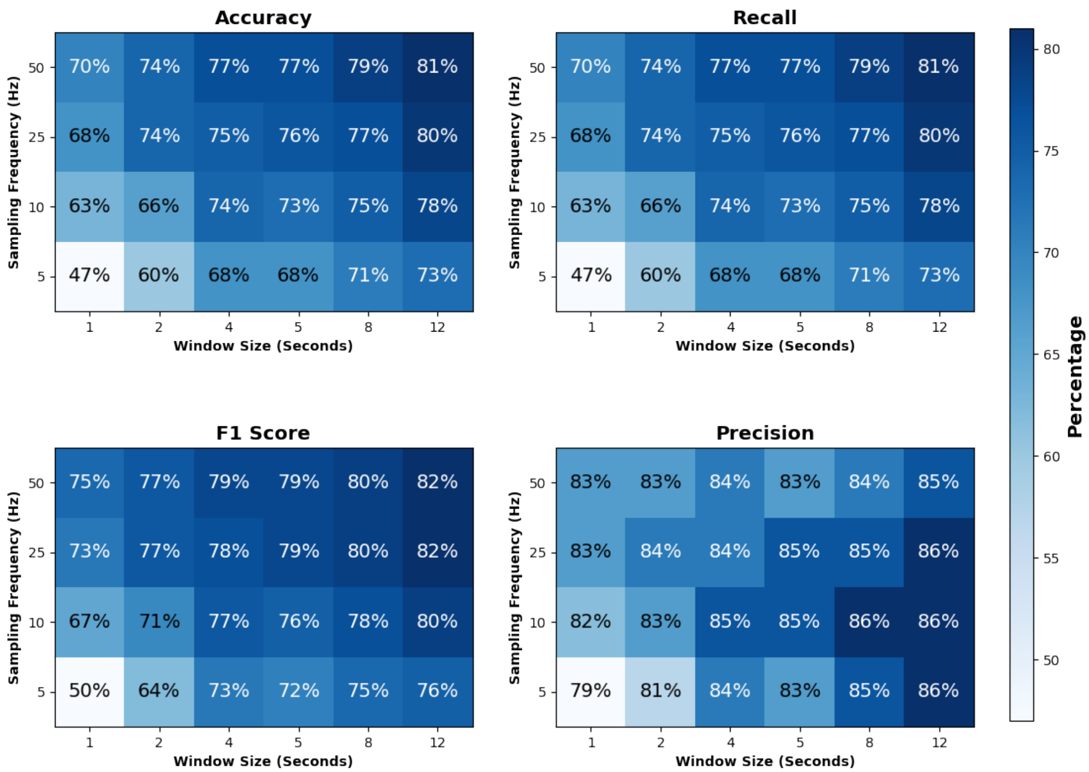

Figure 8. Heatmap of Different Metric vs. Window Size and Sampling Rate.

while still leveraging the advantages of larger windows for improved model stability.

These findings are significant for applications with limited computational resources or where real-time processing is critical. The selection of window size and sampling rate should be tailored to the specific application, taking into account computational resources, real-time constraints, and the nature of the activity data. For real-time monitoring systems, a balance between window size and sampling rate is necessary to ensure timely and accurate classification results.

4.2.3. 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 data or grouping. The goal is to isolate the effects of window size and sampling rate on performance metrics—accuracy, precision, recall, and F1-score.

A mixed-effect model was used to account for both fixed effects and random effects. In this context, the predictors of interest—window size (1s, 2s, 4s, 5s, 8s, 12s) and sampling rate (5 Hz, 10 Hz, 25 Hz, 50 Hz)—were treated as fixed effects, as they directly impact classification performance. To control for individual subject variability, a random effect was included for the 23 subjects in the study. This ensures that differences across subjects do not drive the observed trends but rather serve as random noise, allowing findings to generalize across participants.

The results, summarized in Table 2, show the impact of window size and sampling rate on the performance metrics relative to the baseline condition (12s window size and 50 Hz sampling rate for F1-score). Negative coefficients suggest a decrease in metric value compared to the baseline, while positive coefficients indicate an improvement.

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

A clear trend emerges, demonstrating that reducing the window size from 12s to shorter durations (1s, 2s, 4s, 5s, 8s) results in a significant decrease in accuracy. This finding aligns with the expectation that shorter window sizes may capture less information, leading to lower classification performance. Similarly, lowering the sampling rate from 50 Hz to 5 Hz or 10 Hz significantly reduces accuracy. A higher sampling rate captures more data points, likely contributing to better model performance.

This analysis confirms that the observed trends in classification performance are primarily driven by window size and sampling rate, not by individual subject differences. By accounting for subject variability, the mixed-effects model strengthens the validity of the findings.

4.3. Resting Vitals Analysis

In this section, the relationship between the rolling 3D magnitude measure and deviations in heart rate from a subject's resting values was examined . Understanding this connection is crucial, as it informs how physical activity influence cardiovascular responses and contributes to designing more effective health monitoring systems..

To explore this relationship, heart rate deviation was computed for all data points during activity periods by subtracting each subject's median resting heart rate (determined from 'lying' and 'sitting' periods before walking) from their observed heart rate. A Pearson correlation analysis was conducted between the rolling 3D magnitude and heart rate deviation to assess the strength and significance of their relationship.

Figure 9 presents a scatter plot of rolling 3D magnitude versus heart rate deviation. The analysis reveals a statistically significant correlation (Pearson $r = 0.29$, $p < 0.0001$), suggesting that greater physical activity is associated with larger deviations from the resting heart rate.

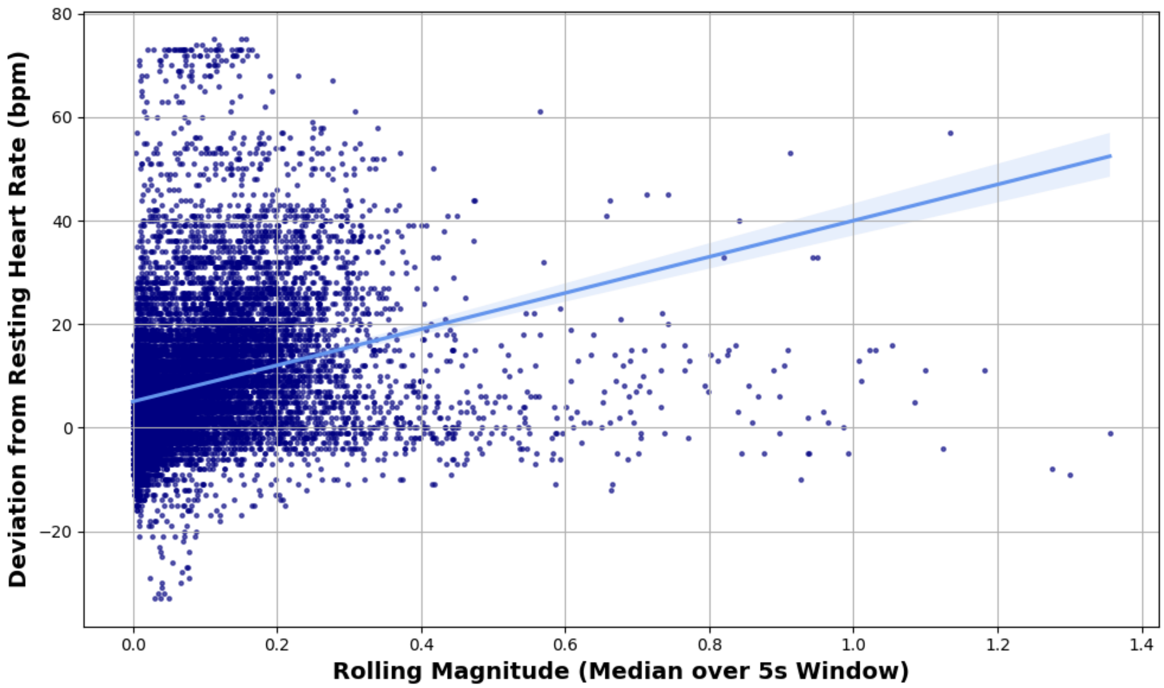

Figure 9. Relationship between Activity Count and Heart Rate Deviation.

5. 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 capability is particularly relevant for conditions like cardiovascular disease and congestive heart failure, where physical activity plays a critical role in disease management. By refining real-time activity classification, accelerometry-based systems can enhance patient care by providing continuous, non-invasive monitoring in real-world environments.

The combined use of 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 machine learning model, particularly the CNN, enables more detailed and precise activity classification, allowing for a more nuanced understanding of patient behavior

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), we were able to quantify deviations in heart rate due to movement. 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, in real-time cardiovascular monitoring, 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.

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. These complementary methods allow for both real-time monitoring and more nuanced behavioral assessments, which are critical for personalized health interventions.

Despite these promising findings, practical challenges remain in balancing model complexity and computational efficiency. The choice of window size and sampling rate significantly impacts model performance. Longer windows and higher sampling rates generally increase accuracy, but also require greater computational resources, making real-time processing more challenging. Future work should explore hybrid models that balance these trade-offs while maintaining robust classification performance.

Additionally, 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. Expanding the dataset to encompass a broader range of populations and real-world scenarios will also improve generalizability and clinical applicability.

All code and data used in this research has been made publicly available Nikookar et al. (2025).

6. 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 can help contextualize physiological measurements, minimizing motion-induced inaccuracies and enhancing 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 linked to conditions such as over-sedation, delirium, or sleep-related illnesses. The proliferation of low cost and low-energy accelerometer devices provides coupled with advanced activity recognition techniques, can provide critical context for biomedical parameters, leading to 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

Ajerla, D., Mahfuz, S. and Zulkernine, F. (2019). A real-time patient monitoring framework for fall detection, *Wireless Communications and Mobile Computing* **2019**(1): 9507938.

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.

Bonato, P. (2003). Wearable sensors/systems and their impact on biomedical engineering, *IEEE Engineering in Medicine and Biology Magazine* **22**(3): 18–20.

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.

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.

Kiarashi, Y., Suresha, P. B., Rad, A. B., Reyna, M. A., Anderson, C., Foster, J., Lantz, J., Villavicencio, T., Hamlin, T. and Clifford, G. D. (2024). Off-body sleep analysis for predicting adverse behavior in individuals with autism spectrum disorder, *IEEE Journal of Biomedical and Health Informatics* .

Kuster, R. P., Grooten, 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.

Maher, C., Szeto, K. and Arnold, J. (2021). The use of accelerometer-based wearable activity monitors in clinical settings: current practice, barriers, enablers, and future opportunities, *BMC Health Services Research* **21**: 1–12.

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.

Poncette, A.-S., Mosch, L., Spies, C., Schmieding, M., Schiefenhövel, F., Krampe, H. and Balzer, F. (2020). Improvements in patient monitoring in the intensive care unit: survey study, *Journal of Medical Internet Research* **22**(6): e19091.

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.

Sathyanarayana, S., Satzoda, R. K., Sathyanarayana, 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., Boppre, 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.