











Raising the Bar(Ometer): Identifying a User's Stair and Lift Usage Through Wearable Sensor Data Analysis

Hrishikesh Balkrishna Karande ,
Ravikiran Arasur Thippeswamy Shivalingappa , Abdelhafid Nassim Yaici ,
Iman Haghbin , Niravkumar Bavadiya , Robin Burchard  ,
and Kristof Van Laerhoven 

University of Siegen, 57076 Siegen, Germany
robin.burchard@uni-siegen.de
<https://ubicomp.eti.uni-siegen.de/>

Abstract. Many users are confronted multiple times daily with the choice of whether to take the stairs or the elevator. Whereas taking the stairs could be beneficial for cardiovascular health and wellness, taking the elevator might be more convenient but it also consumes energy. By precisely tracking and boosting users' stairs and elevator usage through their wearable, users might gain health insights and motivation, encouraging a healthy lifestyle and lowering the risk of sedentary-related health problems. This research describes a new exploratory dataset, to examine the patterns and behaviors related to using stairs and lifts. We collected data from 20 participants while climbing and descending stairs and taking a lift in a variety of scenarios. The aim is to provide insights and demonstrate the practicality of using wearable sensor data for such a scenario. Our collected dataset was used to train and test a Random Forest machine learning model, and the results show that our method is highly accurate at classifying stair and lift operations with an accuracy of 87.61% and a multi-class weighted F1-score of 87.56% over 8-second time windows. Furthermore, we investigate the effect of various types of sensors and data attributes on the model's performance. Our findings show that combining inertial and pressure sensors yields a viable solution for real-time activity detection.

Keywords: Real-time activity recognition · Wearable inertial sensing · Barometric pressure sensing · Stairs and lift taking detection

1 Introduction

The proliferation of wearable technology in recent years has significantly transformed how we monitor and understand human physical activities. Wearable devices, equipped with a variety of sensors, have emerged as pivotal tools in numerous fields including health monitoring, fitness tracking, and medical diagnostics [10, 14]. Activity recognition, a key application of wearable technology,

involves the identification and classification of various physical actions performed by an individual, such as walking, running, sitting, or in our case more complex activities like climbing stairs or using an elevator [4, 16].

Differentiating between activities such as stair climbing and lift use is especially important since each activity places different physiological and biomechanical demands on the body [12]. Stair climbing is a physically challenging activity that can yield useful information on cardiovascular health, lower body strength, and overall mobility. It is also connected with a high caloric expenditure, making it an important exercise to track in fitness and weight control settings [3, 6]. In contrast, using an elevator indicates stationary behavior, emphasizing times of immobility that are critical to understanding overall activity patterns and such behavior [18].

The use of pressure sensors in wearable electronics provides a precise and effective approach for capturing the finer details of these actions. Pressure sensors detect differences in pressure at different positions or in various situations, providing extensive information on an individual's gait and posture [8]. For example, the pattern of pressure changes while ascending stairs differs significantly from the pressure change while standing still in an elevator. Using these sensors, we can improve the accuracy of activity recognition, allowing us to distinguish between the energy demands of stair climbing and the passive nature of the elevator use.

The ability to accurately recognize and differentiate various activities offers profound implications across multiple domains. In personalized healthcare, it enables tailored interventions and monitoring, enhancing treatment plans and preventive care [24]. Additionally, In the context of elderly care, detecting the specific activity of stair climbing versus elevator use becomes critical in ensuring safety. Many older adults are at a higher risk of falls when ascending or descending stairs, and timely detection of stair use can help caregivers monitor and prevent potential accidents or offer immediate assistance when needed [7]. In this way, monitoring specific activities can contribute to safer and more secure living environments for vulnerable populations. In the fitness industry, these technologies allow for more customized exercise plans and real-time feedback on performance, aiding in achieving personal fitness goals [20]. Furthermore, In workplace ergonomics and occupational health, recognizing whether employees are using stairs or elevators can assist in designing more efficient movement patterns within the workplace. Employers can encourage healthier habits such as stair climbing, which can reduce sedentary behavior, improve cardiovascular health, and enhance overall well-being among workers [9]. Additionally, analyzing activity patterns can help assess whether employees are adhering to ergonomic safety protocols, potentially reducing the risk of workplace injuries associated with improper movements or excessive inactivity.

Our dataset¹ and this paper's preliminary experiments offer several contributions:

¹ Both the code and data (raw and resampled form) for the project are publicly accessible on GitHub: https://github.com/iiMox/project_work_stairs_lift_detection.git.

- Introduction of a stairs and lift taking dataset: We present a dataset specifically dedicated to distinguishing between lift usage and stair climbing with multiple sensors and a high variability between participants.
- Characterization of barometric pressure sensor data: We provide a proof of concept for the application of pressure sensor data in activity recognition.
- Examination of Window Size for Time-Series Data: We investigate the impact of various window sizes on the pre-processing of time-series data.
- Feature Importance Analysis: We analyze the importance of different features, highlighting the critical role of pressure measurements in accurately classifying activities.

2 Related Technologies

2.1 Sensors

The field of activity recognition using wearable technology has seen significant advancements due to the integration of various sensor technologies and data analysis techniques [25]. In our study, we utilized accelerometer and barometer sensors to capture the necessary data for activity recognition. The accelerometer measures the acceleration forces acting on the device, providing detailed information about movement and orientation. The barometer, on the other hand, measures atmospheric pressure, which can be useful in detecting changes in elevation, such as when a person uses stairs or an elevator [15].

2.2 State of the Art and Limitations

The development and validation of activity recognition systems have heavily relied on various publicly available datasets. These datasets contain sensor data collected from wearable devices, which were placed on participant’s bodies and recorded during different physical activities. The datasets are crucial for training and evaluating machine learning models. This section reviews some of the most commonly used datasets in the field that include stair usage, highlighting their contributions.

Among the prominent datasets, the UCI HAR [2] dataset includes activities such as “walking upstairs” and “walking downstairs” performed by participants wearing smartphones on their waists. Similarly, the PAMAP2 [17] dataset captures a range of activities including “ascending stairs” and “descending stairs” using multiple sensors placed on participants’ bodies. The MHealth dataset [5] also includes “climbing stairs” along with other physical exercises. Table 1 lists some of the public activity recognition datasets that contain “stairs” as one of their activities. The table also includes information about the recorded sensors and the overall size of the datasets mentioned.

None of the datasets we examined that included stair usage involved the activity of taking lifts. Furthermore, to our knowledge, there are no datasets that specifically involve the activity of using lifts. There are also devices like “Fitbit”

that can detect if you are taking the stairs [19], using sensors such as altimeters to measure elevation changes. However, these devices do not focus on differentiating between the usage of lifts and stairs specifically, and user reports [1, 21] indicate that they cannot reliably detect the difference between lifts and stairs.

In addition to these datasets, other related works pursue the same goal of promoting healthy living through wearable technology. For instance, Neves et al. [11] explore the use of wireless sensor networks to monitor health metrics and encourage healthier behaviors. Similarly, Ohtaki et al. [13] developed a system that automatically classifies ambulatory movements and estimates energy consumption using accelerometers and a barometer. Although these studies do not focus on differentiating between lift and stair usage, they contribute to the broader objective of using wearable devices and sensor technologies to enhance health by monitoring physical activity and energy expenditure, aligning closely with our study's focus on leveraging sensor technology for health promotion.

Our study, therefore, aims to fill this gap by providing extensive sensor data from participants specifically using lifts and stairs.

Our focus is on capturing the use of lifts and stairs, and not just having incidental usage. Our study, therefore, aims to fill this gap by providing extensive sensor data from a larger cohort of participants specifically using lifts and stairs. This level of detail and focus on lift versus stair usage is not comprehensively covered by existing datasets, making our publicly available dataset the first with this focus.

3 Data Collection and Analysis

A total of 20 participants were involved in the data collection process. These participants varied in age (26.0 ± 10.75), gender (10 male and 10 female), and physical fitness levels to ensure a diverse dataset that reflects different human movement patterns. Prior to participation, each individual signed a consent form, acknowledging their voluntary involvement and understanding of the experiment's procedures and objectives. Additionally, we submitted a detailed proposal to The Council for Ethics in Research (Ethics Council) of the University of Siegen, which reviewed and approved the study, allowing us to proceed with the experiment. The data was gathered using the Bangle.js 2 smartwatch, which is equipped with both accelerometer and barometer sensors. The accelerometer measures the acceleration forces acting on the device in three-dimensional space, while the barometer provides data on atmospheric pressure, which can be used to infer altitude changes.

3.1 Data-Collection Procedure

The experiment was conducted in a university building that comprises a total of eight floors that can be switched between by either flights of stairs or a set of elevators. This environment was chosen due to its accessibility and the availability of both stairs and elevators. To ensure variability and reduce systematic bias, we employed randomization in our experiment in several ways:

Table 1. State-of-the-art datasets that are related to activity recognition and contain going up or down the stairs as an activity. The datasets differ in the contained activities, numbers of participants, and sensors used, as well as in their sizes. Our newly created dataset is the only available dataset that includes taking the lift as an activity and uses a barometric pressure sensor.

Dataset	Activities	Stairs	Lift	Users	Sensors used	Dataset size
UCI HAR (Human Activity Recognition Using Smartphones) [2]	Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, Laying	Yes	No	30	Accelerometer, Gyroscope	10,299 Observations (Each 2.56 s)
PAMAP2 (Physical Activity Monitoring) [17]	18 total - basic, household, and exercise activities	Yes	No	9	Accelerometer, Gyroscope, Magnetometer, Heart Rate Monitoring	About 10 h (IMU data: 100 Hz Heart rate data: 9 Hz)
WISDM [22, 23]	18 total - ambulation-related activities, hand-based activities of daily living, and various eating activities	Yes	No	51	Accelerometer, Gyroscope	2754 min
MHealth [5]	12 total - Basic, Locomotion, and exercise activities	Yes	No	10	Accelerometer, Gyroscope, Magnetometer, Electrocardiogram (heart monitoring)	Not specified
Ours	Walking Upstairs, Walking Downstairs, Lift up, Lift down, Null	Yes	Yes	20	Accelerometer, Barometer	525.02 min (50 Hz)

- The next floor for each participant to visit was determined by a random number generator, ensuring an unpredictable sequence of floor visits. The floor from which the experiment would be started was also determined by the random number generator. Since the university building that we used has floors numbered from 2 (first accessible floor) to 8 (last accessible floor), our random number generator had been set to produce the lowest possible random number 2 and the highest possible random number 8.
- The choice between using the elevator or stairs for each transition was decided by a coin toss, ensuring an impartial probability of selecting either mode of movement.
- The participants were advised to act normal and try different variations of activities that they would ideally do while performing the movement of taking the stairs or taking the Lifts. Some participants chose to carry a handbag or a set of books with them, while others used their phones while taking the lifts or stairs. The participants were encouraged to do the experiment at their own pace and with their own choice of effort level. These variations of the different hand movements that the participants did while participating in the experiment were recorded using the accelerometer.

Each participant was instructed to wear the Bangle.js 2 smartwatch throughout the experiment. They were guided to move between floors based on the outcomes of the random number generator and the coin toss. The smartwatch continuously recorded accelerometer and barometer data during these activities at a frequency of 50 Hz. For each participant, the experiment was conducted over a period of approximately 30 min. Participants were given adequate breaks between movements to ensure they were not fatigued, which could affect the naturalness of their movements. The data collection sessions were scheduled to avoid peak hours in the building, minimizing external disruptions and ensuring safety.

Each participant was accompanied by four researchers, each one of these four were responsible for one of these tasks:

1. Video recording the participant's complete physique for the entire duration of the experiment.
2. Annotating the changes in the events that are taking place, classifying them into the 5 classes, and also noting down the time when this change took place.
3. Generating a random number to decide which floor to go to next.
4. Tossing the coin to decide whether to take the stairs or the lift to this floor.

During the data annotation phase, using the video recordings and based on the context data from the annotation data files, class labels were manually added to the sensor data CSV (Comma-Separated Values) files as ground truth for the experiment.

3.2 Activity Classes

The dataset was categorized into five distinct classes:

- **Null**: Periods in which none of the target activities was performed, includes idleness (such as sitting, standing still, walking, or pacing).
- **Lift Up**: Movement of participants using the elevator to ascend floors.
- **Lift Down**: Movement of participants using the elevator to descend floors.
- **Stairs Up**: Movement of participants using the stairs to ascend floors.
- **Stairs Down**: Movement of participants using the stairs to descend floors.

We gathered a total of 525 min (8:45 h) of data from the participants. The distribution of this data among the five classes can be seen in Fig. 1. The exact duration of each class in minutes is given in Appendix D Table 5 (Figs. 2, 3, 4 and 5).

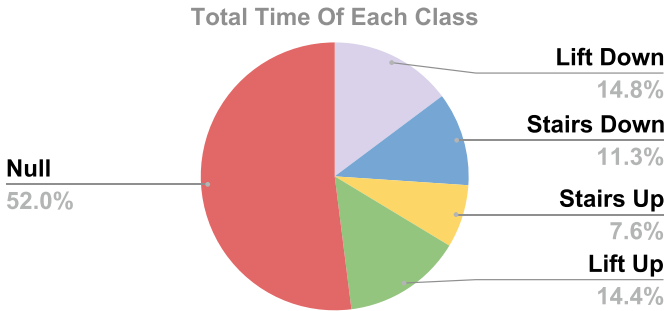


Fig. 1. The distribution of the activity classes contained in the recorded dataset. The Null class makes up 52%, while the rest of the data consists of the four target classes (Lift Up, Lift Down, Stairs Up, Stairs Down).

3.3 Analysis

Sensor Data: The sensor data contains detailed information about the movement and pressure measurements taken at various time intervals. This data is evaluated to better comprehend movement patterns, acceleration variations, and possible activities or behaviors.

- Time: The timestamp indicating the time at which the data was recorded.
- Timestamp: A numerical representation of time, in milliseconds.
- X, Y, Z: Accelerometer data representing the acceleration along the X, Y, and Z axes, respectively.
- Magnitude: Magnitude of the acceleration vector.
- Pressure: Pressure data.
- Label: Ground truth label for the data. One of: Stairs Up, Stairs down, Lift up, Lift down & Null.

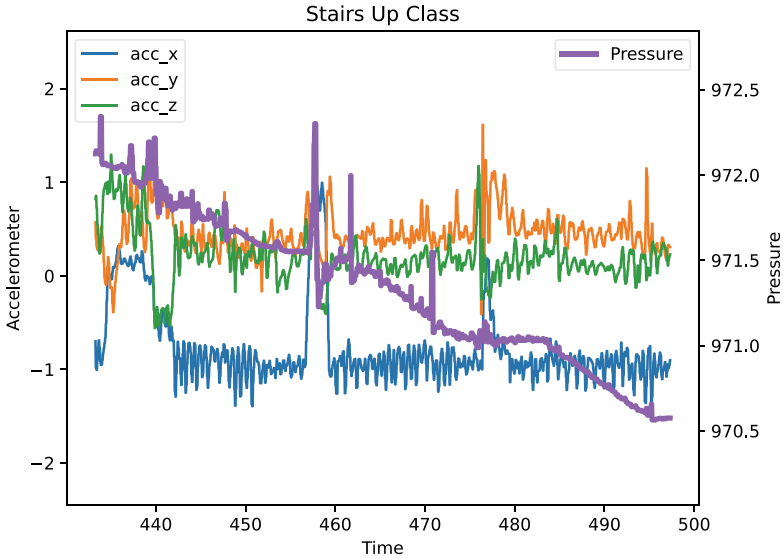


Fig. 2. Example accelerometer and barometer data for the ‘Stairs Up’ activity class. The accelerometer shows the typical patterns of walking. Meanwhile, the pressure steadily decreases, with short plateaus occurring on the stairway’s landings. Additionally, the pressure sensor shows some spikes which are likely caused by the movement of the arm while walking.

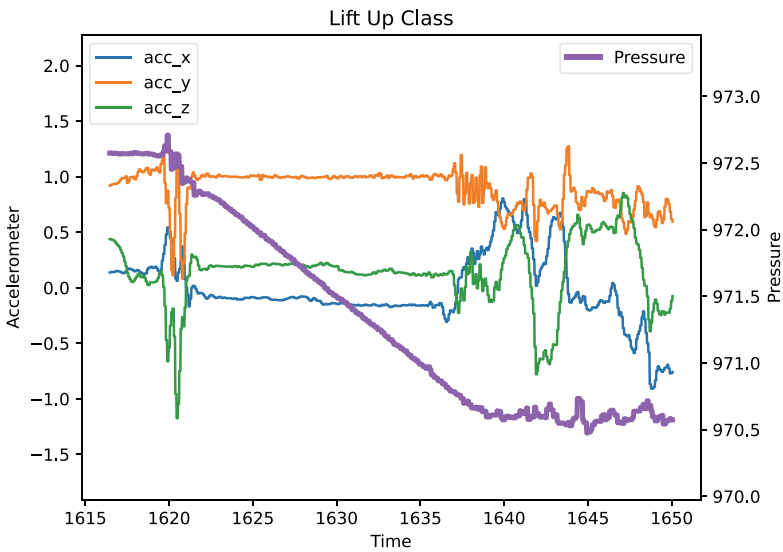


Fig. 3. Example accelerometer and barometer data for the ‘Lift Up’ activity class. While the lift is moving, the participant mostly stood still. We observe the pressure decreasing smoothly.

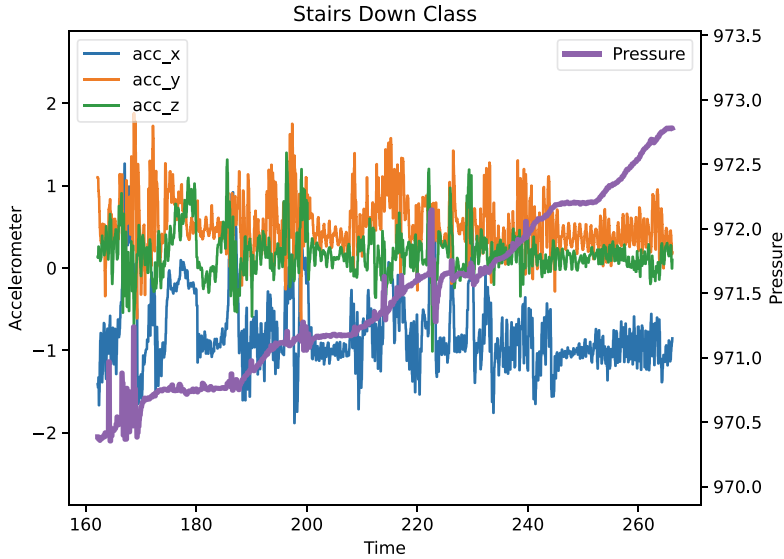


Fig. 4. Example accelerometer and barometer data for the ‘Stairs Down’ activity class. The accelerometer shows a lot of activity, with similar walking patterns as in ‘Stairs Up’. In direct comparison, the magnitude and frequency of the movement sensors seem higher. The barometer detects the steadily rising pressure, also interrupted by short plateaus caused by the landings.

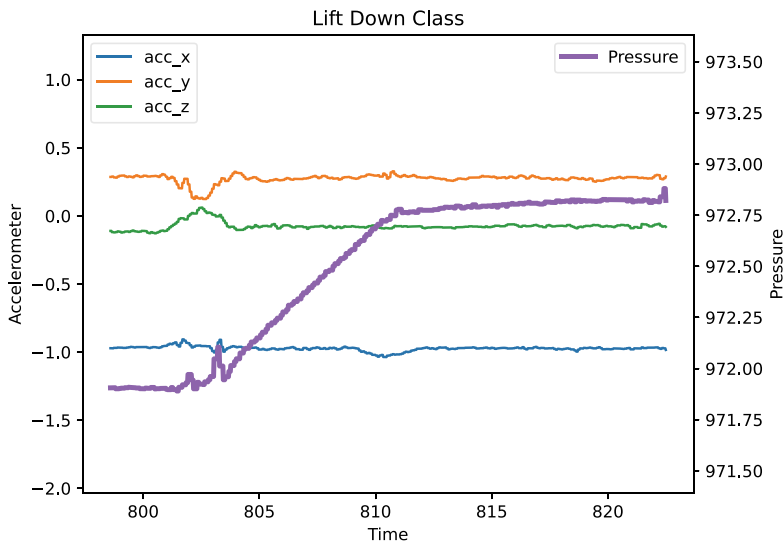


Fig. 5. Example accelerometer and barometer data for the ‘Lift Down’ activity class. The specific participant for this instance was standing very still during the descent. We can see the pressure increasing linearly and smoothly.

Annotation Data: The annotation data adds context to sensor data by designating specific time points as events or activities. This annotation can help segment or categorize sensor data based on various activities.

- Elapsedtime: The elapsed time since the start of the recording.
- Comment: Annotations or labels provided at specific time points, such as “Lift down”, “Stairs up”, etc.

During the initial study of the sensor data, it was discovered that the annotation for “Lift” occurrences was completely dependent on the user entering the lift. However, it became clear that this approach did not accurately represent the user’s experience because the lift may remain stationary for an extended amount of time before moving. This difference resulted in an erroneous portrayal of lift utilization in the data.

To address this issue and ensure the lift data’s accuracy, it was decided to manually annotate certain data points within the sensor data that corresponded to lift occurrences. This manual annotation entailed carefully studying the sensor data and determining the exact instant when the lift began moving.

4 Dataset Validation and Prediction Algorithm

When constructing a detection algorithm for evaluating sensor data, both the data separation method and the detection algorithm selection were carefully considered. To evaluate the collected dataset, a random forest classifier (from `scikit-learn.ensemble` libraries) has been chosen as the classification algorithm for this problem. Before feeding the raw data into the random forest classifier, we are pre-processing the data to extract features using sliding windows. We trained and evaluated the model on different window lengths of 4 and 8 s. For each of the windows, we are extracting 26 features from the data and the label is based on majority voting. During the window creation, windows are discarded if no label is assigned to at least 80% of the total number of observations within the window. We also discard windows that are too short or are missing data, e.g. at the end of a recording.

We use leave-one-participant-out cross-validation to get a subject-independent performance evaluation. In the prediction algorithm, the first step involves selecting a participant to serve as the test participant, while the remaining 19 participants are used for training the model. For these 19 participants, various features are computed as outlined in the data pre-processing stage. These features include statistical measures such as averages, minimums, maximums, variances, and standard deviations for accelerometer data (X , Y , Z), magnitude, and pressure, as well as additional metrics like range, slope, kurtosis, and skewness.

The class imbalance was addressed using the random oversampling technique, which duplicates randomly selected samples from the minority classes until all classes have the same number of samples. The *imblearn* library has been used to oversample the data with ‘not majority’ as the sampling strategy. This step

is meant to prevent the model from being biased towards the majority class and to ensure it performs well across all classes. For this proof-of-concept study, we did not explore the use of undersampling or a combination of oversampling and undersampling. A more sophisticated sampling technique like SMOTE could improve our model’s performance further, but was out of the scope of this work.

To optimize the random forest classifier’s performance, a grid search was conducted to identify the best hyperparameters, specifically the depth of the trees (`max_depth`) and the number of trees in the forest (`n_estimators`). Grid search involves testing various combinations of these hyperparameters to find the combination that yields the highest performance, in our case based on 10-fold cross-validation results. We used ‘GridSearchCV’ from the scikit-learn library. With the optimal hyperparameters identified, the random forest model was trained using the 19 participants in each LOSO split’s training data.

Finally, for each left-out participant, the trained Random Forest model was then used to make predictions and evaluate its performance. The model’s performance was evaluated by calculating the accuracy of its predictions, which measures the proportion of correctly classified instances out of the total number of instances. Additionally, we report the F1-score in three modes of multiclass-averaging (Micro, Macro, Weighted).

To measure the impact of the addition of the pressure sensor to the data, we performed a small ablation study, repeating the same training and validation procedure, but without the pressure data. This enabled us to test the hypothesis, that pressure data contains viable information to detect stairs and lift usage, but also to distinguish between these two, as well as their directions (up or down).

5 Results and Discussion

We show the results of our machine learning experiments in Table 2. The results demonstrate significant performance improvements when the time-series data is pre-processed using an 8-second window for feature extraction compared to a 4-second window. The model’s evaluation metrics, including accuracy and F1 scores, are notably higher with the 8-second window, reaching a macro F1 score of 0.86 (8 s) versus 0.76 (4 s). This substantial increase in accuracy highlights the importance of selecting an appropriate window size for pre-processing time-series data, as it can greatly enhance the classifier’s ability to understand patterns and make accurate predictions.

When comparing the performance of a Random Forest model trained solely on acceleration data (IMU data) to one that incorporates both acceleration and pressure data, the results clearly show that the inclusion of pressure data significantly enhances the model’s ability to accurately distinguish between stair use and lift use. Without the barometric sensor, a macro F1-score of only 0.49 was reached. While acceleration data captures the dynamic movement patterns, pressure data can offer insights into altitude changes, such as those experienced when ascending or descending stairs. The altitude changes experienced when ascending or descending stairs are distinct from the relatively constant altitude

changes during elevator use. For the evaluation metrics of each participant in both time-windows please refer to our GitHub repository.

Table 2. The model evaluation metrics averaged over all 20 participants for 8-second and 4-second time-window. The results are shown with and without the use of the recorded pressure data. The columns on the left show results when IMU and the pressure sensor were used, while the columns on the right show the results without the pressure data.

Sensors	IMU & Press.		IMU only	
	8 s	4 s	8 s	4 s
Accuracy	0.88	0.80	0.68	0.65
F1-Score (Micro-Avg)	0.88	0.80	0.68	0.65
F1-Score (Macro-Avg)	0.86	0.76	0.49	0.46
F1-Score (Weighted)	0.88	0.80	0.67	0.63

Evaluating F1 scores in micro, macro, and weighted modes offers a comprehensive understanding of the model’s performance across different aspects of the data, especially when data is imbalanced. The macro F1 score is particularly informative in scenarios with class imbalance because it reveals the model’s performance across each class individually. The weighted F1 score provides a balanced view that accounts for class imbalance while still reflecting the overall performance across all classes. The table of complete metrics for each participant in 8-second and 4-second time windows along with the best parameter values for *max_depth* and *n_estimators* hyperparameters while fitting random forest classifier can be found in Appendix B. 3 and C.4 respectively.

Figures 6 and Appendix A Fig. 7 present the confusion matrix of the model over an 8-second and a 4-second time window respectively. The confusion matrix offers a detailed breakdown of the model’s performance by presenting the counts of true positives, true negatives, false positives, and false negatives for each class. In Fig. 6, we can observe that most prediction mistakes are related to the null class and that the classes related to taking stairs and taking the lift are rarely confounded.

In the feature importance analysis conducted across all 26 features for each of the 20 participants, it was observed that the feature “slope_pressure” consistently exhibited the highest importance score. This finding underscores the significance of pressure values in effectively distinguishing between the five distinct classes. The steepness or gradient of pressure changes, as represented by the slope feature, appears to be particularly informative in differentiating between activities involving changes in elevation, such as ascending or descending stairs and using lifts. The feature importance scores of all 26 features are presented in the Appendix E as Table 6.

Overall, the evaluation of the collected dataset with the explained methods shows that lift and stair usage can be distinguished and differentiated from the

True Label	Null	1633	10	19	70	61
	lift down	30	175	0	30	0
	lift up	36	3	177	0	9
	stairs down	83	1	0	562	10
	stairs up	101	0	0	15	609
	Predicted Label	Null	lift down	lift up	stairs down	stairs up

Fig. 6. Confusion Matrix for 8-second windows. Most predictions are correct (diagonal of the matrix). We observe that most mistakes revolve around the null class and that ‘stairs’ and ‘lift’ can be separated reliably by the model.

background class with a sufficiently high accuracy and F1 score, given the set of wearable sensors.

6 Conclusions

This paper presents the possibility of differentiating movement into five different classes with acceleration and pressure readings from a smartwatch. The feature importance score analysis depicts the importance of pressure measurement in this particular classification problem. Barometric measuring is a sensing modality seldom used in HAR. This work serves as a proof of concept, highlighting the potential of utilizing pressure data in activity recognition. The findings indicate that pressure-related features, particularly the slope of pressure changes, are highly informative for detecting and classifying different movement patterns related to altitude changes. This insight could pave the way for more advanced and accurate activity recognition systems, which have applications in various applications such as health monitoring, fitness tracking, and smart home environments.

Furthermore, the effectiveness of different window sizes for pre-processing time-series data in the context of activity recognition using a random forest classifier is also investigated. Our findings highlight once more the critical role

of window size selection in enhancing model performance. This underscores the importance of capturing more comprehensive temporal information to better discriminate between activity classes.

We believe the main contributions of our dataset and this paper’s experiments are relevant to the Activity Recognition community:

- A dataset dedicated specifically to taking lifts versus taking stairs detection was collected and annotated
- Our experiment serves as a proof of concept for the use of barometric air pressure data in Activity Recognition
- We have investigated especially the window size in such time-series data pre-processing
- We analyzed Feature Importance scores to investigate the impact of both sensors and features

7 Future Work and Enhancements

The dataset used in this study presents opportunities for further investigation, particularly in addressing imbalances among the five activity classes. Despite our efforts to collect data in an unbiased manner, the inherent variability in participants’ activities has led to class imbalances. The lifts are faster than the stairs, which is one of the main reasons for the imbalance in data. In longer and real-world settings, the data is likely to be even more imbalanced. Future studies could explore strategies to mitigate this imbalance and to deal with it both during data collection and real-time inference.

Expanding the dataset size is another avenue for future research. Currently, data collection occurs within a single building, potentially limiting the generalizability of the model. By collecting data from diverse environments and settings, including outdoor and indoor scenarios, we can enhance the model’s ability to generalize across different contexts. Additionally, increasing the dataset size by recruiting participants from various demographics and activity levels can enrich the diversity and representativeness of the dataset. With a larger dataset, implementing neural networks such as Long Short-Term Memory (LSTM) or transformer-based networks becomes feasible, allowing for the exploration of deep learning architectures capable of capturing even more complex temporal dependencies.

Furthermore, the definition of the Null class can be refined to incorporate a broader range of activities. While the null class primarily represents periods of inactivity, such as standing or sitting, future studies could consider including additional activities like cycling, jogging, or walking within this class. This expansion would provide a more comprehensive representation of daily activities, thereby improving the model’s performance in real-world settings where activities may be more diverse and dynamic.

Acknowledgements. This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 425868829 and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

A Confusion Matrix for 4-Seconds Time-Window

True Label	Null	3340	39	54	216	175
	lift down	147	304	4	9	1
	lift up	173	15	321	1	9
	stairs down	236	0	0	1061	71
	stairs up	281	0	0	89	1153
		Null	lift down	lift up	stairs down	stairs up
		Predicted Label				

Fig. 7. Confusion Matrix for 4-second windows.

B Evaluation Metrics for 20 Participants in 8-Seconds Time-Window

The following table presents evaluation metrics and also suitable hyperparameter values for random forest found by GridSearchCV for each participant in 8-second windows.

Table 3. The model evaluation metrics (Acc for Accuracy, Est.s for Estimators) for all 20 participants for 8-second time windows along with hyperparameter values.

No.	Acc.	macro F1	weighted F1	Depth	Est.s
1	0.84	0.83	0.84	None	200
2	0.85	0.82	0.85	None	250
3	0.90	0.89	0.90	None	250
4	0.82	0.81	0.83	15	275
5	0.96	0.94	0.96	15	250
6	0.87	0.86	0.87	20	200
7	0.81	0.79	0.81	15	225
8	0.89	0.80	0.88	20	200
9	0.90	0.87	0.90	None	275
10	0.91	0.85	0.90	20	250
11	0.92	0.91	0.92	None	300
12	0.90	0.90	0.90	20	200
13	0.87	0.87	0.86	20	300
14	0.89	0.88	0.89	None	250
15	0.87	0.87	0.87	15	350
16	0.88	0.89	0.88	None	300
17	0.86	0.82	0.86	None	300
18	0.88	0.89	0.88	20	300
19	0.86	0.79	0.86	20	275
20	0.85	0.84	0.85	20	225

C Evaluation Metrics for 20 Participants in 4-Seconds Time-Window

The following table presents evaluation metrics and also suitable hyperparameter values for random forest found by GridSearchCV for each participant in 4-second windows.

Table 4. The model evaluation metrics (Acc for Accuracy, Est.s for Estimators) for all 20 participants for 4-second time windows along with hyperparameter values.

No.	Acc.	macro F1	weighted F1	Depth	Est.s
1	0.77	0.72	0.77	None	350
2	0.81	0.75	0.80	None	350
3	0.81	0.76	0.81	None	250
4	0.77	0.73	0.78	None	225
5	0.78	0.73	0.78	None	200
6	0.79	0.76	0.79	None	200
7	0.75	0.72	0.76	None	250
8	0.78	0.75	0.77	None	350
9	0.85	0.80	0.84	None	250
10	0.84	0.70	0.83	None	350
11	0.87	0.70	0.83	None	350
11	0.87	0.87	0.87	None	325
12	0.80	0.78	0.80	None	325
13	0.80	0.74	0.79	None	250
14	0.80	0.72	0.80	None	325
15	0.83	0.82	0.83	None	275
16	0.81	0.79	0.81	None	300
17	0.79	0.74	0.79	None	225
18	0.77	0.77	0.77	None	275
19	0.83	0.78	0.83	None	275
20	0.77	0.73	0.76	None	200

D Total Time Duration for Each Class in Minutes

Table 5. Duration of data collected for each class in minutes

class	Minutes
Lift Down	77.60
Stairs Down	59.20
Stairs Up	39.88
Lift Up	75.36
Null	272.98

In total 152.95 min of data was collected for the lift classes and 99.08 min of data was collected for the stairs classes.

E Feature-Scores for 26 Features Averaged over 20 Participants in 8-Seconds and 4-Seconds Time-Window

Table 6. Feature scores for all features extracted over 8-seconds and 4-seconds window

Feature	8-s	4-s
slope_pressure	0.313075423	0.24633301
var_magnitude	0.084037245	0.079129592
std_magnitude	0.082081804	0.072645593
max_magnitude	0.045221187	0.051749068
std_pressure	0.043930271	0.030623624
var_pressure	0.043276678	0.031141125
min_magnitude	0.036980905	0.032534554
kurtosis_pressure	0.036261681	0.029589955
min_accX	0.027751388	0.028021951
range_pressure	0.02538081	0.025415817
std_accX	0.023843761	0.034228149
var_accX	0.023416437	0.036206354
std_accZ	0.020009935	0.02788877
var_accZ	0.018040161	0.026410733
avg_accX	0.016969519	0.02231106
max_accX	0.016947775	0.022569319
avg_magnitude	0.0169168	0.027153292
skew_pressure	0.015893485	0.019383114
min_accZ	0.015800581	0.020498954
max_accY	0.015392096	0.020042188
std_accY	0.014100437	0.021778705
var_accY	0.01405132	0.021217197
avg_accZ	0.013379868	0.019155533
min_accY	0.013372234	0.017988043
avg_accY	0.012822567	0.017824296
max_accZ	0.011045632	0.018160005

References

1. Jovin, I.: Fitbit floors climbed not accurate? here's what to do. (2019). <https://gadgetsandwearables.com/2019/07/28/fitbit-floor-count-too-high/>. Accessed 26 June 2024
2. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L., et al.: A public domain dataset for human activity recognition using smartphones. In: Esann, vol. 3, p. 3 (2013)

3. Atlas: Cardiovascular benefits of stair climbing: Fitness explained. <https://atlasbars.com/blogs/fitness-explained/cardiovascular-benefits-of-stair-climbing-fitness-explained>. Accessed 26 June 2024
4. Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., Amirat, Y.: Physical human activity recognition using wearable sensors. *Sensors* **15**(12), 31314–31338 (2015)
5. Banos, O., et al.: mHealthDroid: a novel framework for agile development of mobile health applications. In: Pecchia, L., Chen, L.L., Nugent, C., Bravo, J. (eds.) IWAAL 2014. LNCS, vol. 8868, pp. 91–98. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13105-4_14
6. Captain calculator: Calories burned on stairs | calculator & formula (2020). <https://captaincalculator.com/health/calorie/calories-burned-stairs-calculator/>. Accessed 26 June 2024
7. Delahoz, Y.S., Labrador, M.A.: Survey on fall detection and fall prevention using wearable and external sensors. *Sensors* **14**(10), 19806–19842 (2014)
8. Kim, Y., Oh, J.H.: Recent progress in pressure sensors for wearable electronics: from design to applications. *Appl. Sci.* **10**(18), 6403 (2020)
9. Kritzler, M., Bäckman, M., Tenfält, A., Michahelles, F.: Wearable technology as a solution for workplace safety. In: Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, pp. 213–217 (2015)
10. Mattmann, C., Amft, O., Harms, H., Troster, G., Clemens, F.: Recognizing upper body postures using textile strain sensors. In: 2007 11th IEEE International Symposium on Wearable Computers, pp. 29–36. IEEE (2007)
11. Neves, P., Stachyra, M., Rodrigues, J.: Application of wireless sensor networks to healthcare promotion. *J. Commun. Softw. Syst.* **4**(3), 181–190 (2008)
12. Blackmer, N.: Here’s how many stairs you should climb a day for a healthy heart (2023), <https://www.verywellhealth.com/daily-stairs-for-a-healthy-heart-8349369>. Accessed 26 June 2024
13. Ohtaki, Y., Susumago, M., Suzuki, A., Sagawa, K., Nagatomi, R., Inooka, H.: Automatic classification of ambulatory movements and evaluation of energy consumptions utilizing accelerometers and a barometer. *Microsyst. Technol.* **11**, 1034–1040 (2005)
14. Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M.: A review of wearable sensors and systems with application in rehabilitation. *J. Neuroeng. Rehabil.* **9**, 1–17 (2012)
15. Ramasamy Ramamurthy, S., Roy, N.: Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisc. Rev. Data Mining Knowl. Discov.* **8**(4), e1254 (2018)
16. Randell, C., Muller, H.: Context awareness by analysing accelerometer data. In: Digest of Papers. Fourth International Symposium on Wearable Computers, pp. 175–176. IEEE (2000)
17. Reiss, A., Stricker, D.: Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th International Symposium on Wearable Computers, pp. 108–109 (2012). <https://doi.org/10.1109/ISWC.2012.13>
18. Rozalynn S. Frazier, C.P.T.: Do you have a sedentary lifestyle? Here are 8 signs and solutions (2024). <https://www.realsimple.com/health/fitness-exercise/sedentary-lifestyle-signs>. Accessed 26 June 2024
19. TheTechyLife: Which fitbits count stairs (2024). <https://thetechylife.com/which-fitbits-count-stairs/>. Accessed 20 June 2024
20. Thompson, W.R.: Worldwide survey of fitness trends for 2020. *ACSM’s Health Fitness J.* **23**(6), 10–18 (2019)

21. Nielson, T.: [fix] fitbit not counting floors climbed or tracking incorrectly (2023). <https://wearholic.com/fitbit-not-counting-floors/>. Accessed 26 June 2024
22. Weiss, G.: Wisdm smartphone and smartwatch activity and biometrics dataset data set. UCI Machine Learning Repository (2019). <https://archive.ics.uci.edu/ml/datasets/WISDM+ Smartphone+ and+ Smartwatch+ Activity+ and+ Biometrics+ Dataset+>. Accessed 18 May 2022
23. Weiss, G.: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset. UCI Machine Learning Repository (2019)
24. Wu, M., Luo, J.: Wearable technology applications in healthcare: a literature review. *Online J. Nurs. Inform* **23**(3) (2019)
25. Zhang, S., et al.: Deep learning in human activity recognition with wearable sensors: a review on advances. *Sensors* **22**(4), 1476 (2022)