Human activity recognition models for the prediction of daily activities in older patients

DAVID ABRAHAM GONZALEZ NARANJO

June 2023

Abstract: The increase in the older population presents a significant challenge to healthcare systems worldwide, highlighting the imperative for the development of effective monitoring and support systems that can enhance the quality of life for older adults. As older individuals face age-related physical and cognitive limitations, predicting and understanding their daily activities becomes crucial for delivering personalized care, promoting their independence and quality of life. Human Activity Recognition (HAR) plays a crucial role in various domains, including healthcare. It involves the development of machine learning models that help predict human activities based on sensor data collected from various sources. Our findings carry practical implications, contributing valuable insights to the field of HAR. The practical utility of our research lies in its ability to guide the development and implementation of activity recognition systems tailored to the unique needs of older patients. By employing these effective approaches in real-world healthcare contexts, the quality of life and well-being of older individuals can be significantly enhanced, fostering a more comprehensive and personalized care experience. We trained three different baseline machine learning models for HAR, namely K-Nearest Neighbors (k-NN), Support Vector Machine (SVM), RF (Random Forest). With the Random Forest achieving the best results with an F1-score of 0.906.

Keywords: physical activity behavior; human activity recognition; public dataset; machine learning; accelerometer

1 Introduction

People worldwide are living longer, so it is only natural that in the present days, the aging population is becoming one of the world's primary concerns. By 2030 the share of the population aged 60 years and over will increase from 1 billion in 2020 to 1.4 billion. By 2050, the world's population of people aged 60 years and older will double (2.1 billion). The number of persons aged 80 years or older is expected to triple between 2020 and 2050 to reach 426 million (World Health Organization, 2022). This substantial increase will have significant social and health care consequences.

Aging is accompanied by physiological changes, functional decline, and an increased susceptibility to health issues among older individuals (Tari et al., 2019; Danneskiold-Samsøe et al., 2009). As a result, healthcare providers face the challenge of delivering personalized care that addresses the unique needs of this growing population.

The ability to predict and understand the daily activities of older patients is of great importance for providing effective caregiving, tailored care, mitigating risks, promoting their independence, and ensuring their overall well-being. However, the diverse nature and intricate patterns of these activities pose significant challenges for healthcare providers. In order to examine the influence of physical activity on different health outcomes, it is essential to employ objective and reliable tools for assessing physical behavior in everyday life. The utilization of wearable sensors, such as accelerometers, for activity monitoring has emerged as the preferred approach and is progressively being employed in population research studies to complement or replace self-reported measures of physical activity (Wijndaele et al., 2015). To monitor physical, functional, and cognitive health of older adults in their home, HAR is emerging as a powerful tool (Wang et al., 2019), leveraging advanced sensing technologies and machine learning algorithms to automatically recognize and classify activities performed by older patients.

Human activity recognition (HAR) is the process of automatically recognizing and categorizing human activities based on sensor data. HAR has gained significant attention in recent years due to the advancements in sensor technologies and the increasing popularity of wearable devices such as smartphones or smartwatches. In the context of healthcare, HAR has the potential to enable personalized health monitoring, behavior analysis, and intelligent systems that adapt to human needs, as it can provide valuable tools for emergencies. Investigating HAR can be approached using different methods and data sources such as accelerometer signals, smartphones or smartwatches. Body-worn accelerometers are the most commonly used data collection method to support HAR due to their low cost and small size (Demrozi et al., 2020).
Several studies have trained and tested machine learning models on self recorded datasets but only a few are publicly available. Additionally, some machine learning studies for accelerometer-based HAR were performed in a controlled space specifically designed for the scientific experiments but the data shows that machine learning models developed in laboratory conditions demonstrate poor performance when tested outside the laboratory (Gyllensten and Bonomi, 2011), thus the importance of performing the experiments in free-living conditions, where participants are free to perform activities of their everyday life, providing more valuable insights into human behavior and activity patterns in real-world contexts.
The aim of this work was to find accelerometer-based datasets where the experiments were conducted under free living conditions. **Falta poner los resultados de nuestra investigación aquí**

2 Related work

2.1 HAR approaches

Only a limited number of HAR research papers explore the utilization of multiple accelerometers, even though classification performance can be improved if doing so (Olguín and Pentland, 2006; Cleland et al., 2013). For instance, Stewart et al. (2018) trained a Random Forest (RF) classifier using a laboratory-recorded dataset from 75 participants (42 children, 33 adults) who wore two Axivity AX3 accelerometers on their thighs and lower backs. This study achieved a balanced accuracy of 99.1% for adults and 97.3% for children in predicting six activities: sitting, lying, standing, slow walking, fast walking, and running. Similarly, Narayanan et al. (2019) recorded free-living data from 30 participants (15 children, 15 adults) who wore AX3 accelerometers on their thighs, lower backs, and wrists. The combination of thigh and lower back sensors yielded the best balanced accuracy of 95.6% (adults) and 92% (children) using an RF classifier. Bao and Intille (2004) explored the use of up to five bi-axial accelerometers (right hip, dominant wrist, non-dominant upper arm, dominant ankle, non-dominant thigh) worn by 20 subjects performing 20 activities. Their investigation involved four classifiers, with the decision tree achieving the highest accuracy of 84%. Olguín and Pentland (2006) examined acceleration data collected from up to three sensors (wrist, hip, chest). While using all three sensors yielded the best accuracy at 92.1%, using a combination of wrist and hip sensors still achieved similar results at 87.2%. Hip/wrist configurations were also studied by Ahmadi et al. (2020), where a RF classifier was trained on free-living data of preschool-aged children's activities. The combination of hip and wrist accelerometers outperformed the individual sensors for specific activities. Baños et al. (2012) investigated the use of nine sensors (one for each limb and upper back) and trained a k-NN classifier (achieving the best performance), a decision tree, and a nearest class center classifier. Ugulino et al. (2012) employed an AdaBoost classifier and four accelerometers to classify five activities, achieving an overall weighted accuracy of 99.4%. Zubair et al. (2016) used the same dataset as Ugulino et al. (2012) to train RF and AdaBoost classifiers, with RF outperforming AdaBoost with an overall accuracy of 99.9% and an averaged precision and recall of 99.8%.
While multiple works have utilized more than two sensors, Olguín and Pentland (2006) argue that the inclusion of additional sensors does not significantly enhance HAR results. Furthermore, a lower number of sensors can provide greater participant comfort during data collection.
In addition to these studies, Logacjov et al. (2021) developed the Human Activity Recognition Trondheim Dataset (HARTH) where 22 participants performed different activities during their regular working hours while carrying out their everyday activities as naturally as possible. Two experts annotated twelve activities in total. They used two accelerometers placed on the thigh and lower back to collect sensor data. HARTH provides high-quality acceleration measurements with fixed sensor placements and professionally annotated labels. They compared the performance of 7 supervised machine learning approaches for HAR, namely

K-Nearest Neighbors (k-NN), Support Vector Machine (SVM), RF (Random Forest), Convolutional Neural Network (CNN), bidirectional LSTM, extreme gradient boost (XGB), and CNN with multi-resolution modules (multi-resolution CNN). With the best model being SVM.
Ustad et al. (2023) conducted a comparable study known as: The Human Activity Recognition 70+ (HAR70+), where 18 older adults aged between 70 and 95 years were included in the data collection. The semi-structured validation protocol took place in and nearby the participants' own homes. They were equipped with two accelerometers and a chest-mounted camera. They reproduced the work of Bach et al. (2021) by training the XGB approach on the HARTH dataset, then the HARTH and the HAR70+ datasets were combined to train a second XGB.

3 Methods

3.1 Human Activity Recognition Trondheim Dataset

Logacjov et al. (2021) provided a number of CSV files, containing readings of two tri-axial Axivity AX3 accelerometers for data acquisition. The AX3 is a small (23 × 32.5 × 7.6 mm) and lightweight (11 g) sensor. The reasons they used two sensors are the following. Cleland et al. (2013) investigated up to six sensors but observed no significant increase in performance compared to two sensors. Hence, two sensors provide high accuracy, higher comfort for the participants, and reduced costs (Reiss and Stricker, 2012). The term "HARTH" is the abbreviation for "Human Activity Recognition Trondheim." It is named after the place it was recorded.
For the sensor positions, one sensor was attached to each participant's right, front thigh (approximately 10 cm above the upper kneecap), and the other to their lower back (approximately 3rd lumbar vertebra). Seen from the participant's perspective while standing upright, the lower back sensor's x-axis points downward, the y-axis to the left, and the z-axis forward. For the thigh sensor, the y-axis points to the right and the z-axis backward.
A video camera was placed on each participant's chest using a chest harness, pointing downwards to record leg movements, later used for annotation. Twenty-two healthy adults (eight female) were recruited via word of mouth between university and hospital staff. They were on average 38.6 ± 14 years old (range: 25–68), had an average height of 177.3 ± 8.3 (range: 157–191) cm, an average weight of 72.9 ± 10.6 (range: 56.0–92.0) kg, and an average BMI of 23.1 ± 2.3 (range: 19.2–28.4) kg/m2.
The dataset was recorded in two sessions. In the first session, 15 participants (six female) were instructed to perform their everyday activities while two sensors recorded acceleration data. The participants performed activities such as sitting, standing, lying, walking, running, and additional activities like stairs, shuffling, cycling, and transportation. The total recording time in the first session was approximately 30 hours, with an average duration of 120 minutes per participant. Videos of the activities were annotated frame-by-frame using a coding scheme.
Due to imbalances in the class labels in the first session, a second data collection session was conducted to primarily focus on walking, running, and cycling. Participants performed activities in various settings, including flat, uphill, and downhill sections. The second session resulted in approximately 7 hours of recorded data, with an average duration of 60 minutes per participant. The data was annotated by human experts using the ANVIL annotation tool. Despite the addition of the second session, the dataset still exhibited imbalances in the activity labels, which poses challenges for training reliable machine learning models.

3.2 Human Activity Recognition 70+

To be included, participants had to be aged 70 years or older and be able to walk independently with or without walking aids. Eighteen older adults aged between 70 and 95 years were included in the data collection. Four participants used a walker during all walking activities and one participant used walking sticks when walking outdoors. Participants were fitted with two Axivity AX3 accelerometers positioned on the lower back and the right thigh. A GoPro Hero 8 camera was attached to the chest of the participants with a chest harness. The camera was pointing downwards filming only the abdomen and lower limbs of the participants. A semi-structured free-living protocol for classification of physical behavior was carried out. The protocol included the activity types walking, standing, sitting, and lying. The first part of the validation protocol took place inside the home of the participants (approximately 25 min), and the second part took place outside (approximately 15 min). The researcher focused on keeping the behaviors as natural as possible to avoid a "laboratory feeling" and encouraged the participants to change between activity types using simple instructions. The video analysis was used as the ground truth for the validation of the activity types identified by the ML model.

3.3 Preprocessing

The preprocessing steps performed before training the machine learning models are described in the following section. First files were extracted, then with each file, time- and frequency-domain features were extracted out of each window. Due to the findings of Banos et al. (2014) the authors segmented the time series into non-overlapping one-second windows (50 samples at 50 Hz), the same approach of segmenting the data into one-second window was chosen for this work. We consider eight signals for feature computation, the six axes (three for each sensor), and each sensor's vector magnitude $\sqrt{x^2 + y^2 + z^2}$. The gravity component features of the human's orientation and movement were separated by computing the gravity and movement component of the raw accelerometer signal. A fourth-order 1 Hz low-pass Butterworth filter was applied to estimate the gravity component, the role of this step is to reduce the noise of the signals and retain valuable data (Chen et al., 2021). The movement component was calculated by subtracting the resulting gravity component from the raw signal. We computed the mean, the median, the standard deviation, the coefficient of variation, the 25th, and 75th percentile, as well as the minimum and maximum for each frame of the gravity components, to get orientation information. For the movement components, we computed the skew, kurtosis, and signal energy, as well as the frequency-domain features frequency-domain magnitudes' mean, frequency-domain magnitudes' standard deviation, dominant frequency, dominant frequency's magnitude, spectral centroid, and total signal power. Then, we computed the axis correlation between all six axes and between the two vector magnitude signals. We also computed the mean across the gravity components of the two sensors. In total, we generated 161 features for each window.
In contrast to the authors, our choice was to use a standard scaler instead of the min-max scaler to help standardize the data set's features onto unit scale with mean=0 and variance=1. This decision was motivated by the importance of employing the correct scaler to simplify the complexity of sample spaces and reduce dimensionalities using the statistical technique Principal Component Analysis (PCA). In the process of selecting the correct number of components we plotted the cumulative explained variance against the number of principal components and we verified the % of variance we wanted to explain in our data. 26 components were selected because of the desire to achieve an explained variance greater than 95%. The target of each machine learning model is to learn the twelve labels of the dataset. Finally, because of the data being separated in a number of CSV files, a Python program was created that concatenates the preprocessed data into a single CSV file.

4 Experiments and results

The experiments are examined as follow. First hyperparameter optimization was performed combined with cross-validation to find reasonable hyperparameters for a Random Forest (RF), a K-Neighbors Classifier (k-NN), as well as a Support Vector Machine (SVM). The hyperparameter optimization with cross-validation, was conducted by selecting two subjects at random from each session of the aforementioned dataset for testing purposes. The remaining 18 subjects were employed for training the models. This cross-validation technique yielded three iterations, with each iteration having a distinct set of subjects in the test set. Each hyperparameter configuration was trained on these three iterations, and the results were averaged for comparative analysis. The average F1-score, encompassing all twelve labels, was chosen as the performance metric of focus due to its greater resilience to class imbalance compared to accuracy (Banos et al., 2014).
The radial basis function was used as the kernel function for the SVM. We investigated the regularization parameter C, with larger values causing a more substantial penalty on wrongly classified samples (best: 10). Furthermore, we explored various γ values, which represent a parameter of the radial basis function. The optimal value was found to be $\frac{1}{N \times \sigma^2}$, where N represents the number of features (N = 26) and σ^2 corresponds to the variance of the training set X.
The bootstrapping technique is employed for the RF classifier. At each node of a decision tree, a subset

of $\sqrt{N}$ features is randomly selected to identify the optimal split. The Gini impurity criterion is utilized to evaluate the quality of a split. Various quantities of decision trees are evaluated during hyperparameter optimization, with the optimal number determined to be 80. Moreover, different minimum samples required to split a node are investigated, with the optimal value found to be 10.
In the case of the k-NN algorithm, different numbers of neighbors were evaluated [2, 4, 6, 8, 10, 12, 30]. Through hyperparameter optimization, the most favorable value of k was determined to be 10.

5 Discussion

The Human Activity Recognition (HAR) system presented in this study demonstrates the potential of utilizing accelerometer data to accurately classify human activities. By extracting and analyzing features from raw accelerometer signals, the system achieves commendable performance in recognizing different activities, paving the way for numerous practical applications in healthcare, sports, and human-computer interaction. Among the evaluated classifiers, the Random Forest model emerges as the most successful algorithm in distinguishing various human activities. The ensemble nature of Random Forest enables it to combine multiple decision trees, reducing overfitting and enhancing the model's generalization ability. Additionally, the classifier's ability to handle high-dimensional datasets and capture complex interactions between features makes it an attractive choice for activity recognition tasks.
The success of the Random Forest model also highlights the importance of feature engineering. By extracting a diverse set of statistical features from the raw accelerometer signals, the system effectively captures relevant information about the temporal and frequency characteristics of human movements. These features act as valuable inputs to the classifiers, facilitating accurate and efficient classification.
Despite the promising results, a significant challenge in HAR lies in dealing with imbalanced classes. In real-world scenarios, certain activities may be underrepresented in the dataset, leading to biased predictions and reduced performance on those activities. Imbalanced classes may cause the model to favor the majority class and neglect the minority classes, hindering the system's ability to generalize to diverse activity patterns. To address this challenge, researchers should explore various strategies to mitigate the impact of imbalanced classes. Class-weighted techniques, where misclassification costs are adjusted based on class prevalence, can help balance the influence of different classes and improve the model's performance on underrepresented activities.

6 Conclusions

In conclusion, this study showcases the successful application of accelerometer-based data for Human Activity Recognition. The Random Forest classifier stands out as the top-performing algorithm, demonstrating its efficiency and accuracy in classifying various human activities. The extensive feature engineering process, capturing both time-domain and frequency-domain information, contributes to the classifier's robustness and ability to recognize diverse activity patterns.
However, the issue of imbalanced classes remains a critical challenge in the development of reliable HAR systems. To ensure the system's effectiveness and generalization capability, researchers should proactively address the class imbalance problem. Implementing class-weighted techniques, exploring ensemble methods, and utilizing sampling strategies can help mitigate the impact of imbalanced classes and enhance the performance of the HAR system on underrepresented activities.
Looking ahead, further advancements in HAR research can be achieved by incorporating more robust deep learning models, which are capable of automatically extracting relevant features from raw sensor data. The combination of deep learning architectures, feature engineering, and techniques to address class imbalance can yield more sophisticated and accurate HAR systems, driving the field towards more practical and real-world applications. Moreover, continuous efforts in data collection, preprocessing, and model refinement are essential to achieve high-performance HAR systems that cater to the diverse needs of end-users.

References

Ahmadi, M. N., Pavey, T. G., and Trost, S. G. (2020). Machine learning models for classifying physical activity in free-living preschool children. *Sensors*, 20(16).

Bach, K., Kongsvold, A., Bårdstu, H., Bardal, E., Kjærnli, H., Herland, S., Logacjov, A., and Mork, P. (2021). A machine learning classifier for detection of physical activity types and postures during free-living. *Journal for the Measurement of Physical Behaviour*, 5:1–8.

Banos, O., Galvez, J.-M., Damas, M., Pomares, H., and Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, 14:6474–6499.

Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In Ferscha, A. and Mattern, F., editors, *Pervasive Computing*, pages 1–17, Berlin, Heidelberg. Springer Berlin Heidelberg.

Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M., and Amft, O. (2012). A benchmark dataset to evaluate sensor displacement in activity recognition. pages 1026–1035.

Chen, J., Huang, X., Jiang, H., and Miao, X. (2021). Low-cost and device-free human activity recognition based on hierarchical learning model. *Sensors*, 21.

Cleland, I., Kikhia, B., Nugent, C., Boytsov, A., Hallberg, J., Synnes, K., McClean, S., and Finlay, D. (2013). Optimal placement of accelerometers for the detection of everyday activities. *Sensors*, 13(7):9183–9200.

Danneskiold-Samsøe, B., Bartels, E., Bülow, P., Lund, H., Stockmarr, A., Holm, C., Wätjen, I., Appleyard, M., and Bliddal, H. (2009). Isokinetic and isometric muscle strength in a healthy population with special reference to age and gender. *Acta physiologica (Oxford, England)*, 197 Suppl 673:1–68.

Demrozi, F., Pravadelli, G., Bihorac, A., and Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8:210816–210836.

Gyllensten, I. C. and Bonomi, A. G. (2011). Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life. *IEEE Transactions on Biomedical Engineering*, 58(9):2656–2663.

Logacjov, A., Bach, K., Kongsvold, A., Bårdstu, H. B., and Mork, P. J. (2021). Harth: A human activity recognition dataset for machine learning. *Sensors*, 21(23).

Narayanan, A., Stewart, T., and Mackay, L. (2019). A dual-accelerometer system for detecting human movement in a free-living environment. *Medicine & Science in Sports & Exercise*, 52:1.

Olguín, D. and Pentland, S. (2006). Human activity recognition: Accuracy across common locations for wearable sensors.

Reiss, A. and Stricker, D. (2012). Creating and benchmarking a new dataset for physical activity monitoring.

Stewart, T., Hedayatrad, L., Neville, J., Mackay, L., and Duncan, S. (2018). A dual-accelerometer system for classifying physical activity in children and adults. *Medicine & Science in Sports & Exercise*, 50:2595–2602.

Tari, A., Nauman, J., Zisko, N., Skjellegrind, H., Bosnes, I., Bergh, S., Stensvold, D., Selbæk, G., and Wisløff, U. (2019). Temporal changes in cardiorespiratory fitness and risk of dementia incidence and mortality: a population-based prospective cohort study. *The Lancet Public Health*, 4:e565–e574.

Ugulino, W., Cardador, D., Vega, K., Velloso, E., Milidiú, R., and Fuks, H. (2012). Wearable computing: Accelerometers' data classification of body postures and movements. In Barros, L. N., Finger, M., Pozo, A. T., Gimenénez-Lugo, G. A., and Castilho, M., editors, *Advances in Artificial Intelligence - SBIA 2012*, pages 52–61, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ustad, A., Logacjov, A., Trollebo, S. O., Thingstad, P., Vereijken, B., Bach, K., and Maroni, N. S. (2023). Validation of an activity type recognition model classifying daily physical behavior in older adults: The har70+ model. *Sensors*, 23(5).

Wang, Y., Cang, S., and Yu, H. (2019). A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167–190.

Wijndaele, K., Westgate, K., Stephens, S., Blair, S., Bull, F., Chastin, S., Dunstan, D., Ekelund, U., Esliger, D., Freedson, P., Granat, M., Matthews, C., Owen, N., Rowlands, A., Sherar, L., Tremblay, M., Troiano, R., Brage, S., and Healy, G. (2015). Utilization and harmonization of adult accelerometry data: Review and expert consensus. *Medicine and science in sports and exercise*, 47.

World Health Organization (2022). Ageing and health. Available at: `https://www.who.int/news-room/fact-sheets/detail/ageing-and-health`. Accessed: July 4, 2023.

Zubair, M., Song, K., and Yoon, C. (2016). Human activity recognition using wearable accelerometer sensors. In *2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pages 1–5.