Engineering Applications of Artificial Intelligence 151 (2025) 110706



Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai





A Convolutional Neural Network based model for daily motor activities recognition: A contribution for Parkinson's Disease

Helena R. Gonçalves [a], Luís Martins [a], Nuno Ferrete Ribeiro [a], Luís Abreu [a], Ana Margarida Rodrigues [c,d], Cristina P. Santos [a,b,*]

[a] Center for MicroElectroMechanical Systems, University of Minho, 4804 - 533, Guimarães, Portugal
[b] LABBELS - Associate Laboratory, 4800 - 058, Guimarães, Portugal
[c] Neurology Service, Hospital of Braga, 4710-243, Braga, Portugal
[d] Clinical Academic Center, 4710-243, Braga, Portugal

ARTICLE INFO

Keywords:
Parkinson's disease
Motor monitoring
Activities recognition
Deep learning
Convolutional neural network

ABSTRACT

Daily motor activities are affected by motor disabilities caused by Parkinson's disease (PD). Monitoring motor disabilities frequently observed in PD is difficult for physicians, as they are limited to the information observed or self-reported during routine consultations, resulting in subjectivity and limited assessment. Thus, it is necessary to evaluate more frequently and objectively, ideally in a continuous manner including daily life tasks. While wearable sensory devices, such as inertial sensors, and their applications are steadily growing, the advances in artificial intelligence studies have revolutionized their ability to extract deeply hidden information for accurate detection and interpretation of motor activities. However, further studies are required, mainly focused on PD. This study aimed to implement a deep learning (DL) based model for recognizing daily motor activities based on inertial data to contribute to PD. The model relied on a convolutional neural network (CNN) architecture, trained and tested on a created custom dataset. We further benchmarked our model against other popular DL frameworks. The dataset included inertial data captured from 18 patients while performing trivial quotidian tasks, such as walking, turning, sitting, and lying. We hypothesized that a DL model based on a CNN architecture could be an appropriate solution for modeling daily motor non-steady and steady-state tasks from a single inertial sensor. We measured an F1 score of 0.906 and an accuracy of 91.1 % on final testing with our optimized CNN model, being the tasks of standing and walking the most accurately recognized by the model. Future challenges should cover exploring attention-based models and increasing the dataset.

1. Introduction

Difficulties in initiating movements, slowness, loss of movement amplitude, resting tremors, rigidity, postural instability, and freezing of gait are the primary motor disabilities that patients with Parkinson's disease (PD) suffer in day-to-day life (Rovini et al., 2017). Thus, the quotidian trivial motor activities are affected by these motor disabilities and, consequently, patients' quality of life. Monitoring motor disabilities frequently observed in PD is difficult for physicians, as they are limited to the information observed or self-reported during routine consultations, resulting in subjectivity and limited assessment (Dijkstra et al., 2008), (Moore et al., 2011). It is required to increase the frequency of assessment, including more often (out of routine consultations) and ideally continuously and more objectively (avoiding patients' recall of memory) while performing their daily life activities. Continuous motion analysis enables monitoring individualized motor performance changes during the most affected daily activities (Rovini et al., 2017)– (Ye and De, 2017),(Zanardi et al., 2021).

Recognizing daily motor activities can offer physicians a quantitative portrayal of motor function behavior in a natural environment and for an extended duration (Dijkstra et al., 2008), (Moore et al., 2011). As a result, a more impartial evaluation can be established, and subsequently, treatments can be customized to the patients' real needs. Therefore, tracking patients' motor behavior during daily motor activities could constitute a biomarker of illness stage, allowing to adopt personalized treatments and consequently delaying patients' motor symptoms (Ye and De, 2017).

Human activity recognition (HAR) is the process of interpreting

* Corresponding author. Center for MicroElectroMechanical, University of Minho, Azurém Campus, 4800-058, Guimarães, Portugal.
E-mail address: cristina@dei.uminho.pt (C.P. Santos).

https://doi.org/10.1016/j.engappai.2025.110706
Received 30 October 2023; Received in revised form 8 October 2024; Accepted 26 March 2025
Available online 2 April 2025
0952-1976/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

human motion using computational methods. Human motion can be interpreted as activities, gestures, or behaviors recorded by sensors (Vrigkas et al., 2015)– (Baskar et al., 2020). HAR has multifaceted applications due to the multiple applications of wearable sensors (Gupta et al., 2022). While wearable sensory devices and their applications are steadily growing, the advances in Deep Learning (DL) such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) derivatives and other hybrid models have revolutionized the ability to extract deeply hidden information for accurate and continuous recognition and interpretation of motor activities (Vrigkas et al., 2015), (Gupta et al., 2022). CNNs excel in handling spatial and hierarchical features by leveraging convolutional layers to extract local patterns and hierarchical representations from input data (Rav et al., 2017)– (Gil-Martín et al., 2020). On the other hand, Long-Short Term Memory (LSTM) networks, a type of RNN, specialize in capturing long-term dependencies in sequential data (Guan and Plötz, 2017). Furthermore, the Bidirectional LSTM (BiLSTM) approach enhances this capability by incorporating both past and future information, enabling a more comprehensive consideration of long-term temporal contexts (Gil-Martín et al., 2020)– (Martins et al., 2022). Hybrid CNN-LSTM models have gained prominence, merging the strengths of CNNs in spatial feature extraction with the sequential modeling capabilities of LSTM networks, thus offering a powerful framework for tasks requiring both spatial and temporal analysis (Kiprijanovska et al., 2020), (Altuve et al., 2020).

As a result, nowadays, DL is being applied to data captured from wearable sensors (Gupta et al., 2022) for identifying and naming activities. HAR is being applied as a method for identifying and naming activities using DL from data captured with wearable sensors (Gupta et al., 2022). Several investigations have been accomplished to develop new HAR frameworks that provide continuous patient monitoring by combining wearable sensory data with advanced artificial intelligence algorithms (Zwartjes et al., 2010), (Rav et al., 2017), (Kazemimoghadam and Fey, 2022). Inertial sensors have undoubtedly been a promising solution frequently used in the last decades to provide data that feeds artificial intelligence models capable of recognizing steady-state motion, such as sitting, standing up/down, or lying on a bed, or non-steady state motion, like ascending/descending stairs and walking (Rav et al., 2017).

This manuscript introduces two primary innovations in the field of HAR for PD, setting it apart from existing literature (Rav et al., 2017)– (Rueda et al., 2018), (Guan and Plötz, 2017), (Kwon et al., 2014)– (Tahavori et al., 2017) that predominantly focuses on HAR solutions for healthy individuals. Firstly, we present a DL-based model tailored specifically for recognizing daily motor activities in PD patients. To develop this model, we collected data from 18 PD patients performing a range of daily motor tasks, such as walking, turning, sitting, and lying, thus contributing with relevant data. According to the literature (Rav et al., 2017)– (Rueda et al., 2018), we considered that CNNs are particularly effective for modeling both non-steady and steady-state motor. However, we have also explored other prominent DL neural network architectures and frameworks, such as the previously referred LSTM and hybrid CNN-LSTM models.

Secondly, the application of a single inertial sensor for data collection represents another innovative aspect of our work. This choice was driven by the aim to develop an algorithmic solution that leverages a device which can be seamlessly integrated into the daily lives of patients, thereby offering a practical and non-intrusive means for continuous monitoring and assessment of patients' motor profiles.

Overall, our contributions are distinguished by the development of an end-to-end CNN-based model for daily motor activity recognition in PD, supported by a unique dataset of inertial information collected from PD patients. Several ablation techniques were explored as well as LSTM and CNN-LSTM based frameworks for benchmark. This research lays the groundwork for the creation of efficient systems capable of continuous monitoring and provides valuable insights for the performance analysis and prescription of suitable therapeutic interventions for PD patients.

2. Related works

Considering the need to understand the related state-of-the-art, a comprehensive review was accomplished on the scientific contributions of wearable sensors combined with artificial intelligence for HAR. Twelve studies were reviewed, and a description of their context, data input, and algorithms used are provided in Table 1.

Every day, human beings carry out different physical activities. These activities range from the simplest ones, such as walking, sitting, and lying down, to the most complex such as brushing teeth, jumping rope, or vacuuming the room. HAR has a particular impact in the health sector because it allows knowing the habits and health conditions of a patient, thus improving the ability of diagnosis and treatment by health professionals. The literature review found that most authors tried to develop their research on HAR for healthy people. Only one study tried to develop HAR models for people with PD (Tahavori et al., 2017). Thus, it verified that there is still an extensive way to apply these methods to people with mobility disorders.

Most human activities commonly classified are walking, climbing and descending stairs, sitting and standing (Rav et al., 2017)– (Rueda et al., 2018), (Guan and Plötz, 2017), (Kwon et al., 2014)– (Tahavori et al., 2017). Lying is also a slightly less represented activity studied in (Guan and Plötz, 2017), (Kwon et al., 2014), (Xu et al., 2019)– (Inoue et al., 2018), (Huang et al., 2023)– (Cheng et al., 2023). Despite the effort to detect these tasks, the reverse of some tasks, such as getting up from a chair or getting out of bed, has not yet been included. Detecting these activities has a significant clinical value because it characterizes patients' motor performance in different daily activities. Further, most of the selected studies applied HAR in healthy people, being eminent to develop new research for developing new HAR algorithms applied to PD (Rav et al., 2017)– (Rueda et al., 2018), (Guan and Plötz, 2017), (Kwon et al., 2014)– (Tahavori et al., 2017). Thus, developing and customizing HAR systems is required to recognize the basic motor activities of people with PD.

The inertial measurement units (IMUs) were frequently applied to measure users' motion during motor tasks. IMUs are an ideal option considering their low cost, low power consumption, portability, size, weight, accessibility, and usability. Even though most studies collected inertial data with a smartphone, disadvantageously, the captured data from smartphones is hardly integrated with other monitoring or motor assistance devices (Rav et al., 2017)– (Zhang et al., 2022), (Murad and Pyun, 2017)– (Zainudin et al., 2016). The IMUs were used to measure kinematic-driven data from the coupled segment, namely angular velocity, and acceleration. These sensors were mostly placed on the legs, ankles, and trunk, varying in the number of sensors applied. Typically, one sensor was applied when a smartphone was used (Rav et al., 2017)– (Zhang et al., 2022), (Murad and Pyun, 2017)– (Zainudin et al., 2016), while when using the IMUs, multiple sensors were used (Rueda et al., 2018), (Guan and Plötz, 2017), (Xu et al., 2019).

Most of the datasets created are public (Rav et al., 2017), (Garcia et al., 2021), (Zainudin et al., 2016)– (Garcia-Gonzalez et al., 2020), (Zhang et al., 2022), (Rueda et al., 2018), (Guan and Plötz, 2017), (Kwon et al., 2014)– (Inoue et al., 2018). They comprise data collected on healthy people, mostly from smartphones (Garcia-Gonzalez et al., 2020). Even in (Tahavori et al., 2017), whose study purpose was PD, the dataset included only two participants with PD. Using data from a reduced number of subjects may raise problems in the reliability of the artificial intelligence models' performance due to the possibility of overfitting not only in the training process but also in the validation process. Therefore, there is a need to consider using data from more participants, especially patients with PD. Furthermore, there is a need to find a standard to capture the most significant activities to provide information about patients' motion function since all studies used non-common protocols (Gupta et al., 2022).

Table 1
Studies identified in the literature review about computational methods for HAR.

Ref.	Motivation	Activities recognized	Population on dataset	Sensory data input		Artificial intelligence
				Raw data	Device	Model
Rav et al. (2017)	General	Walking, jogging, up/downstairs, sitting and standing	36 Healthy	Angular velocity Acceleration	📱	CNN
Zainudin et al. (2016)	General	Walking, jogging, up/downstairs, sitting and standing	36 Healthy	Angular velocity Acceleration	📱	MLP and LR
Zhang et al. (2022)	General	Walking, jogging, up/downstairs, sitting and standing	36 v	Angular velocity Acceleration	📱	CNN
Inoue et al. (2018)	General	Walking, jogging, skipping, up/downstairs, sitting, standing, and lying	540 Healthy	Angular velocity Acceleration	📱	DT, SVM, RF and RNN
Murad and Pyun (2017)	General	Walking, running, jumping, up/downstairs, sitting, and sleeping	44 Healthy	Angular velocity Acceleration	📱	RNN
Guan and Plötz (2017)	General	Walking, up/downstairs, sitting, standing, and lying	13 Healthy	Angular velocity Acceleration	26 IMUs	LSTM
Xu et al. (2019)	General	Walking, up/downstairs, sitting, standing, and lying	13 Healthy	Angular velocity Acceleration	26 IMUs	RNN
Batool et al. (2019)	General	Walking, jogging, up/downstairs, sitting, standing, and lying	24 Healthy	Angular velocity Acceleration	📱	SVM
Rueda et al. (2018)	General	Walking, jogging, up/downstairs, sitting, standing, and lying	13 Healthy	Angular velocity Acceleration	26 IMUs	CNN
Kwon et al. (2014)	General	Walking, running, sitting, standing, and lying	–	Angular velocity Acceleration	IMU	HIER
Tahavori et al. (2017)	PD	Walking, backward walking, sitting, standing, and lying	10 Healthy + 2 PD patients	Angular velocity Acceleration	IMU	DT, RF and SVM
Garcia et al. (2021)	General	Walking, jogging, up/downstairs, sitting and standing	46 Healthy	Angular velocity Acceleration	📱 & 3 IMUs	EAE
(Huang et al., 2023)– (Cheng et al., 2023)	General	Cycling, walking, rope jumping, lying, sitting on a chair, car driving, playing soccer, watching TV, phone calls, walking, hopping, typing, and waving	9 to 12 Healthy	Angular velocity Acceleration	📱, cameras and 1to 36 IMUs	CNN

Ref.: Reference of the identified studies; **LSTM:** Long Short-Term Memory; **CNN:** Convolutional Neural network; **RNN:** Recurrent Neural Network; **MLP:** Multilayer Perceptron; **LR:** Linear Regression; **DT:** Decision Tree; **SVM:** Support Vector Machine; **RF:** Random Forest; **HIER:** Hierarchical agglomerative clustering; **EAE:** Ensemble Auto Encoder; **H:** Healthy subjects; **PD:** patients with PD; **Min:** minimum; 📱: smartphone; and **IMUs:** Inertial Measurement Units.

Activities classification is essentially a time series problem, and time-series classification is a type of supervised Machine Learning (ML) in the artificial intelligence domain (Gupta et al., 2022). It is used to predict future values from past data using statistical techniques and can be used for forecasting and offloading sensor data. In recent years, due to their popularity, DL technologies, for instance, CNNs and RNNs, have been successfully introduced in HAR applications (Vrigkas et al., 2015), (Gupta et al., 2022), (K et al., 2021). LSTM networks and hybrid CNN-LSTM architectures are gaining attention for HAR tasks. However, their use is primarily limited to public datasets featuring only healthy adult subjects and targeting a narrow range of human activities. Research efforts utilizing these models have demonstrated promising results (Gil-Martín et al., 2020), (Martins et al., 2022)– (Altuve et al., 2020), although some studies have reported performance degradation when validating models in a subject-independent manner (Altuve et al., 2020).

What distinguishes these DL models from the classical statistical ML methods is that they can extract relevant features from raw data (Gupta et al., 2022). Classical statistical ML models, such as support vector machine, random forest, decision trees, or multilayer perceptron classifiers, were trained to identify different activities using processed features (mean, SD, minimum, maximums, or even correlation coefficients). Thus, these approaches require a pre-processing stage that makes difficult real-time implementations (Guan and Plötz, 2017). Indeed, the studies (Zhang et al., 2022) and (Xu et al., 2019) implemented computational methods for HAR in real-time using CNNs and RNNs, respectively. Furthermore, recent studies have highlighted a new trend in HAR research, focusing on fine-tuning large pre-trained models using self-supervised learning, which offers potential for better adaptation to specific HAR problems (Yuan et al., 2024). Thus, it seems reasonable to point out DL models as promising solutions to HAR models with time efficiency requirements or which need low pre-processing phases (Vrigkas et al., 2015), (Gupta et al., 2022). Nevertheless, it is required to adapt the metrics used to evaluate the performance of these DL-based models (Gupta et al., 2022). For example, given the type of data input involved in HAR models present unbalanced datasets (since some activities can take more time than others), the F1 score is a suitable metric to lead with this type of data (Guan and Plötz, 2017), (Xu et al., 2019) instead of using accuracy (Rav et al., 2017)– (Rueda et al., 2018), (Kwon et al., 2014), (Batool et al., 2019), (Murad and Pyun, 2017)– (Zainudin et al., 2016), (Tahavori et al., 2017).

2.1. Highlights

Our review of HAR solutions in the literature identified significant gaps, notably: (i) a shortage of research focused on PD; (ii) datasets predominantly featuring healthy subjects; (iii) unclear guidelines on sensor placement for data collection; (iv) common reliance on smartphones for data collection, which hinders the integration of HAR models into specialized monitoring devices; (v) the absence of a standardized protocol for recording human motor activities; (vi) a deficit in real-time application implementations; and (vii) the use of metrics that do not adequately assess model quality. To bridge these gaps, our study pursued a systematic approach with the goal of develop, validate and

benchmark a DL-based model for recognizing daily motor activities in PD patients using inertial data. Our approach focused on two key areas: (i) developing a solution tailored for PD, utilizing a dataset of motions recorded with a single inertial sensor from PD patients; and (ii) creating this dataset specifically for this purpose, ensuring it was directly relevant and gathered with end-users' needs in mind. This research produced two primary innovations: the model's specificity for PD and the utilization of a unique, purpose-collected dataset from a single sensor.

This model was created for integration with a clinical desktop application. Its function is to automatically detect specific activities and use additional algorithms to calculate motor metrics. These metrics will be supplied to physicians, aiding in the consistent and objective evaluation of patients' motor functions. Given this design, real-time processing was not necessary for our study. Although real-time application was not a requirement in our study, we still considered computational factors to benchmark our optimized CNN-based model with other DL frameworks. This included the number of parameters, memory footprint, training and testing time of each model, allowing us to further explore potential avenues for possible real-time implementations.

3. Methods

3.1. Solution conceptualization overview

Daily motor activities recognition is a time-series classification in which DL-based models based on inertial data have been successfully introduced (Vrigkas et al., 2015), (Gupta et al., 2022). Aligned with this idea, we implemented a DL-based model for daily motor activities recognition in PD. Fig. 1 depicts an overview of the proposed solution.

Motor symptoms in PD usually manifest in everyday situations while performing trivial motor activities. Patients present alterations in gait, slowness, and loss of amplitude in motor transition movements, such as sitting and lying, episodes of akinesia, and FoG events when turning, among others (Rovini et al., 2017). Thus, focusing on the monitoring needs in the PD field, it was defined that the proposed model for daily motor activities recognition should be able to recognize walking, standing, 180deg turning, sitting, getting up from a chair, lying, and getting up from a bed. In contrast to the literature, reverse activities are identified and included (getting up from a chair and a bed).

To ensure the users' freedom of movement and enhance the acceptability of our solution, we chose to employ a single inertial sensor placed at the fifth lumbar (L5) vertebra. This waist-mounted device can capture motion data not only from walking but also from upper body movements. It uniquely facilitates the recording of a complete gait cycle using just one device, which devices attached to the legs cannot achieve. We utilized a pre-validated instrumented waistband, known as sensBand, developed and validated by our team in studies (Guan and Plötz, 2017), (Murad and Pyun, 2017), (Inoue et al., 2018). This development followed a user-driven approach, focusing on system usability. The device's usability was evaluated with 36 patients using the System Usability Scale, achieving a rating (mean ± SD = 77.288 ± 8.21 %) that classifies it as good and acceptable, according to standards (Gonçalves et al., 2021), (Gonçalves et al., 2024). A protocol was delineated to capture inertial data from patients performing the identified activities using the sensBand, which measures 3-axis acceleration and angular velocity via an accelerometer and gyroscope embedded in its inertial measurement unit (IMU), creating the dataset used in the final solution.

DL-based neural networks, specifically CNNs, have proven to be the most effective in performing HAR (Vrigkas et al., 2015), (Gupta et al., 2022). Therefore, we hypothesized that a DL model based on a CNN architecture could be an appropriate solution for modeling daily motor non-steady and steady-state tasks from a single inertial sensor. The DL-based model relied on a CNN architecture with optimized parameters. Thus, the model input was prepared, as well as the relevant metrics to be used to evaluate the model and optimize its performance. After the model's conceptualization, the training and testing phase was followed. Ablation techniques were employed after refining the CNN-based model for optimal performance. The final validated CNN-based model was further evaluated and benchmarked against other popular contemporary architectures, such as LSTM, BiLSTM and several hybrid CNN-LSTM models (Vrigkas et al., 2015), (Gupta et al., 2022), (Romijnders et al., 2023), (Kiprijanovska et al., 2020).

With the proposed DL-based solution, it is expected to provide a tool to characterize patients' motor profiles while performing daily motor tasks. The experimental methodology designed to implement the CNN-based model is depicted in Fig. 2. The model was developed in Python using mainly TensorFlow, Keras, Scikit-learn, Jupyter, and NumPy libraries.

3.2. Data acquisition

A custom dataset was created based on data acquisition conducted under the ethical procedures of the Hospital of Braga Ethical Commission 36/2018, following the Helsinki Declaration and the Oviedo Convention. All participants gave their informed consent to be part of the study. Data were collected at the Hospital of Braga in 2CA. Eighteen participants diagnosed with PD were recruited during their regular visits to the Neurology Service at the Hospital of Braga. Patients were recruited if they presented a score in H&Y lower than four, were aged between 50 and 85 years old, did not have a cognitive impairment, presented an autonomous gait, and were evaluated by the same neurologist. Patients' motor conditions were assessed using the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale part III (MDS-UPDRS III), in which this subcategory is widely used in a clinical setting as well as in research scope to evaluate patients' motor experiences of daily living and motor complications (Goetz et al., 2008). *All patients were able to autonomously perform common daily motor activities (e.g.*, walking, sitting, lying) and

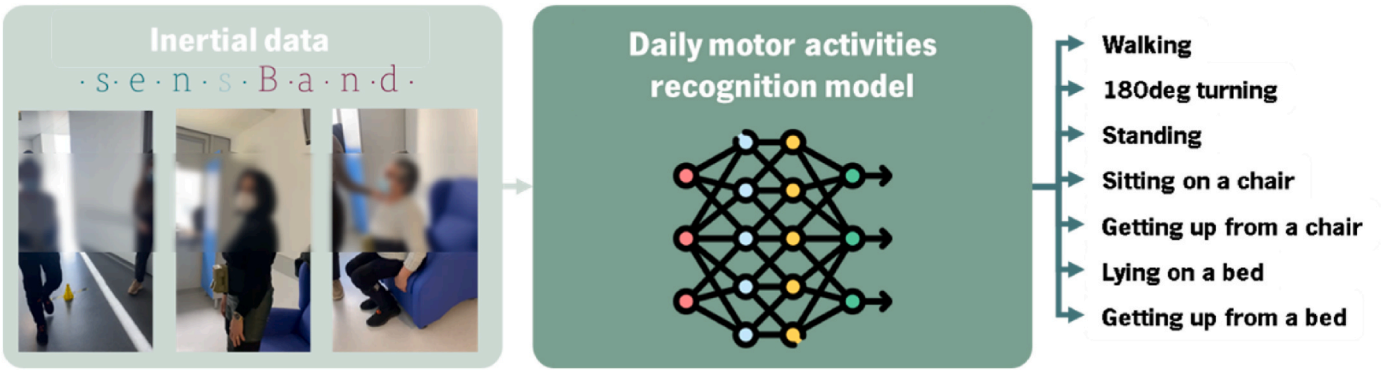


Fig. 1. Overview of the proposed solution for daily motor activities recognition.

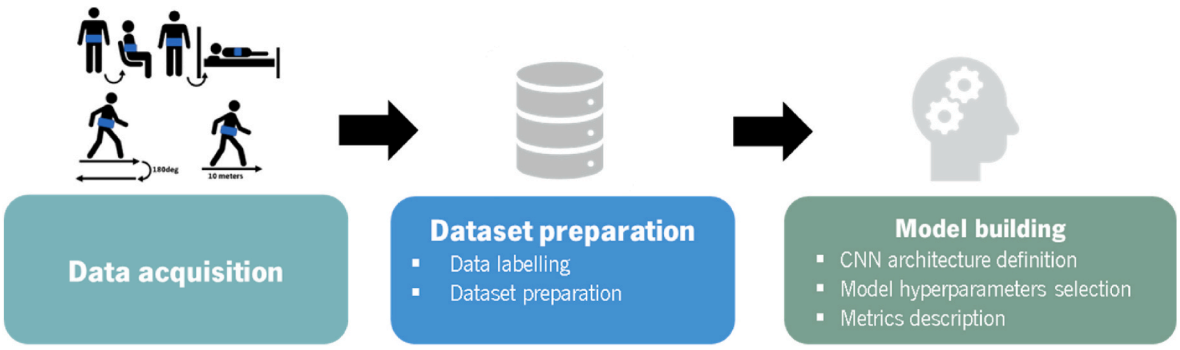


Fig. 2. End-to-end steps to build the CNN-based model for daily activities recognition.

performed the experimental protocol in the ON phase. Table 2 summarizes participants' demographic and clinic characteristics, highlighting the mean and standard deviation (SD) values.

The protocol involved recording inertial data at a sampling rate of 100 Hz using the instrumented waistband (sensBand) shown in Fig. 3, while patients performed the following activities: walking, standing, 180deg turning, sitting, getting up from a chair, lying, and getting up from a bed. Thus, six motor trial activities were delineated and depicted in Fig. 4, to collect data for all labels captured during their execution. Data for each subject was recorded continuously, encompassing three repetitions of the six predefined exercises, resulting in a single file per subject, each clearly identified by an ID. For model classification, we merged all files into a single dataset, ensuring that no windows overlapped between subjects during window generation for training and validation, thus preventing data mixing across individuals. To record the walking activity, the participants stood for 10sec, walked 10m, and stood for 10sec again. Participants walked 10m and performed a 180deg turning to capture turning activity. For collecting data from sitting on a chair, participants stood for 10sec and sat in a chair. The reverse task captured the activity of getting up from a chair. The same procedure was followed to collect the activity of lying on a bed and getting up from the bed. Participants were asked to lie flat in a supine position. The activity of standing was defined as when participants waited to walk, sitting, lying, and getting up/down from a chair/bed. A researcher was responsible for registering the moments when the activity started and ended, using a timer shown in a smartphone APP that started synchronized with the beginning of the sensBand function. The same researcher registered that information in all data acquisitions to avoid bias, do not influence the model performance, and achieve an appropriate process of labeling data.

3.3. Dataset preparation

3.3.1. Data labeling

The data labeling process was manually done based on the synchronized notes of activities start/stop taken during data acquisition. Seven labels were defined: standing, sitting on a chair, getting up from a chair, lying on a bed, getting up from a bed, walking, and 180deg turning, as illustrated in Fig. 5. Each sample captured was labeled with a

Table 2
Sociodemographic and clinic characteristics of study participants.

Characteristics		Participants (N = 18)
Age [years] (mean ± SD)		67.89 ± 8.72
Height [cm] (mean ± SD)		166.22 ± 7.47
Weight [Kg] (mean ± SD)		70.69 ± 11.63
Sex	female	N = 8
	male	N = 10
Hoehn and Yahr scale (mean ± SD)		2.322 ± 1.13
MDS-UPDRS III (mean ± SD)		36.63 ± 24.40



Fig. 3. Instrumented waistband, sensBand, used to collect the inertial data. Focus on the sensory system.

respective scale label ID. As a title of example, Fig. 5 depicts the vertical acceleration in the different activities, and the colors represent the corresponding labels defined. Also, it indicates the used label ID.

3.3.2. Dataset splitting

During the model training and validation phase, we optimized the model using the best hyperparameters identified during a fine-tuning stage, with a grid search cross-validation process using the training dataset. We then divided the training dataset into two parts: 80 % for training and 20 % for validation. The testing dataset, separate from these, was employed exclusively as unseen data for the final evaluation of the model's performance. Therefore, our approach utilized distinct training, validation, and testing datasets. All datasets were normalized using the minimum-maximum method (Zhang et al., 2022).

3.4. Model building and benchmarking

The proposed model was based on a CNN architecture presented in (Brownlee, 2018). The model building pipeline comprised three phases: (1) model hyperparameters optimization through a grid search process with cross-validation in the fine-tuning stage using *training dataset*; (2) model training and validation using the best hyperparameters found in the fine-tuning stage (*training dataset* subdivided with an 80/20 proportion for training/validation); and (3) evaluation of final model performance using unknown data (*testing dataset*). Fig. 6 depicts these stages.

To benchmark our CNN model against other well-known DL architectures, we first applied the optimized hyperparameters, which were determined during the initial phase with the CNN. Then, we followed the steps of model training (phase 2) and model testing (phase 3) for these alternative DL models. This method facilitated a detailed and

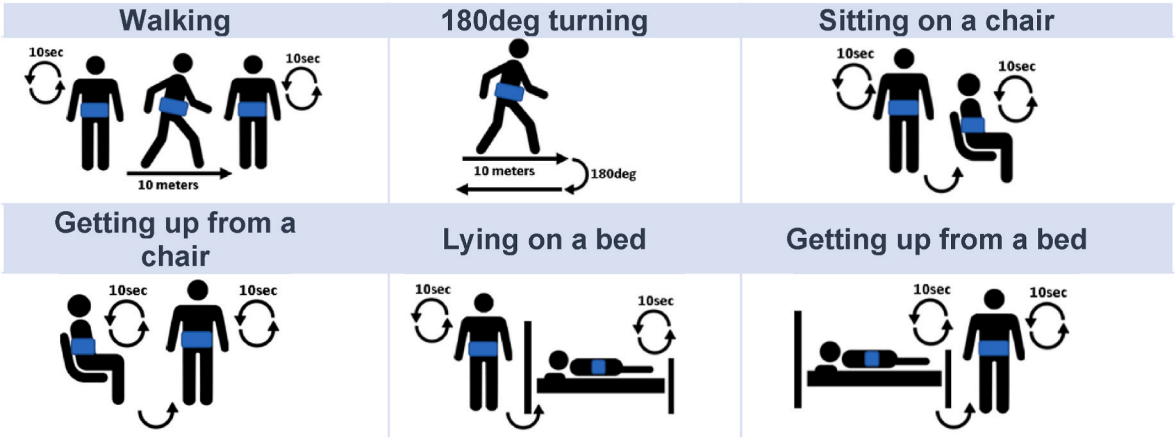


Fig. 4. Experimental procedures to capture data for dataset creation.

consistent comparison across various models, allowing us to evaluate their performance in relation to our primary CNN model, while enhancing the reliability and replicability of our results.

3.4.1. CNN-based model architecture

This research tested several CNN architectures based on (Brownlee, 2018). The proposed CNN architecture model, shown in Fig. 7, is composed of: (1) two 1D convolution layers; (2) a dropout layer; (3) a pooling layer; (4) a flatten layer; (5) a fully-connect layer; and (6) an output layer.

Two convolution layers were defined since, according to (Brownlee, 2018), it gives the model a better chance of learning features from the input data. The dropout layer was intended to help slow down the learning process of the convolution layers, and the pooling layer was added to reduce the learned features to the essential elements. After the convolution and pooling layers, a flatten layer was added. At the end, a fully connected layer was used before the output layer to make a classification. The fully connected layer worked as a buffer between the learned features and the output to interpret the learned features before making a prediction. The output layer was another dense layer with an input shape considering the number of labels.

CNN models usually have three-dimensional input with the following shape: (batch size, window size, number of features). The batch size (N in Fig. 7) corresponds to the number of collected windows that the model runs in each interaction. The window size (timestep in Fig. 7) is the number of samples of each window. Finally, the number of features is the x, y, and z acceleration and angular velocity. Each window is assigned a corresponding label in one-hot-encoding format, which is determined by selecting the most frequent label among all the labels within the window. The output for the model was a seven-element vector containing the probability of a given window belonging to each of the seven activity types. During the fine-tuning phase, different window sizes were experimented to improve model performance. Also, in grid search, other CNN hyperparameters were tested considering the model architecture, such as the number of filters, kernel size, and pooling size. Final model training used the best hyperparameters found in the fine-tuning phase.

3.4.2. Hyperparameters in fine-tuning phase and final model

A set of hyperparameters was defined to test in a grid search, and every different combination of the chosen hyperparameters was tested while calculating the performance metrics using 10-fold cross-validation. These parameters included: the number of filters, kernel size, pooling size, neurons, batch size, drop-rate, and learning rate. Also, different window sizes were experimented, considering the time duration of each motor activity. Further, 50 % of overlap was used based on (Brownlee, 2018) to handle activity transitions more accurately. The experimented values on the different parameters are listed in Table 3, as well as the respective fine-tuned values.

Grid search returns a ranked report of the best results obtained by the different hyperparameter combinations, requiring a scoring metric. Since input data have unbalanced labels, it was chosen to use the F1 score to rank the grid hyperparameters combinations (Guan and Plötz, 2017), (Xu et al., 2019). Moreover, the Adam version of the stochastic gradient descent algorithm was used to optimize the network, and the categorical cross-entropy loss function was applied as this is a multi-class classification problem. Fig. 8 depicts the final model trained with a representation of the layers' input size, considering the best hyperparameters.

3.4.3. Model performance

The precision, recall, accuracy, and F1 score were estimated using equations (1)–(4) to evaluate the model during the training and testing phases. These equations were applied by label and then the performance was presented as a weighted average across all classes. Precision was used to measure the correctly identified positive cases from all the predicted positive cases, while recall was used to estimate the correctly identified positive cases from all the actual positive cases. Accuracy measured all the correctly identified cases. Finally, the F1 score is the harmonic mean of precision and recall, used to measure incorrectly classified cases. Since the proposed solution used an unbalanced dataset, the F1 score was used as the primary measure to analyze the performance model quality to train the suggested model. A confusion matrix was also used to better visualize the model performance for each class classification.

$$Precision_{label\ ID} = \frac{TP_{label\ ID}}{(TP_{label\ ID} + FP_{label\ ID})} \tag{1}$$

$$Recall_{label\ ID} = \frac{TP_{label\ ID}}{(TP_{label\ ID} + FN_{label\ ID})} \tag{2}$$

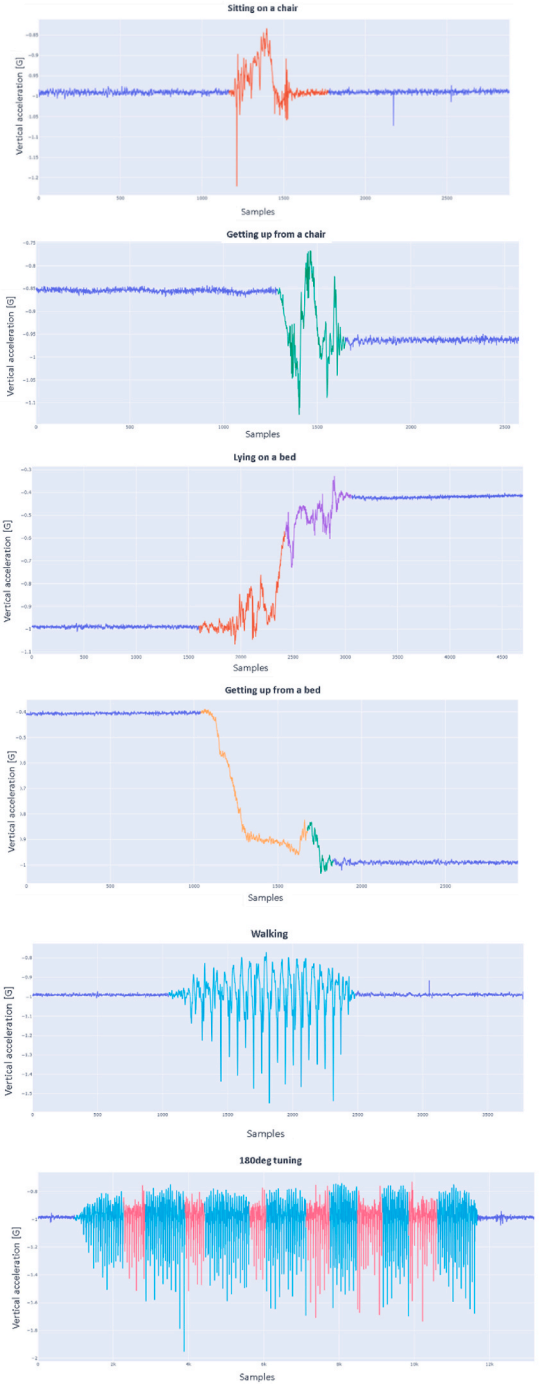
$$Accuracy_{label\ ID} = \frac{TP_{label\ ID} + TN_{label\ ID}}{(TP_{label\ ID} + FP_{label\ ID} + TN_{label\ ID} + FN_{label\ ID})} \tag{3}$$

$$F1\ score_{label\ ID} = 2 \cdot \frac{(Precision_{label\ ID} \cdot Recall_{label\ ID})}{(Precision_{label\ ID} + Recall_{label\ ID})} \tag{4}$$

3.4.4. Ablation model index

An ablation study was conducted to study each layer's contribution to the CNN-based model performance, as indicated in Table 4.

This study consisted of systematically removing layers from the CNN model and verifying the model performance at each removal in order to identify the most suitable candidate for subsequent benchmarking. At each remotion, the ablated model was trained with the *training dataset*



Color	Label ID	Activity
Dark blue	0	Standing
Red	1	Sitting on a chair
Green	2	Getting up from a chair
Purple	3	Lying on a bed
Orange	4	Getting up from a bed
Cyan blue	5	Walking
Pink	6	180deg turning

Fig. 5. Data labeling for each activity represented over the vertical acceleration signal for the different activities with different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and with the same optimized parameters used to build the final model. The results were obtained by testing the new ablated models against the *testing dataset*. Additionally, given the importance of Batch normalization layers for performance and computational cost improvement, a last ablation study was conducted by incorporating Batch Normalization layers along with dropout (Romijnders et al., 2023), (Garbin et al., 2020), (Ioffe and Szegedy, 2015), to study the impact of Batch normalization layers inclusion.

3.4.5. Benchmark deep learning architectures

Building on our main goal of validating a CNN-based model for HAR, we compared it with other alternatives frequently mentioned in HAR research, such as LSTM and combined CNN-LSTM models (Vrigkas et al., 2015), (Gupta et al., 2022), (Romijnders et al., 2023), (Martins et al., 2022), (Kiprijanovska et al., 2020). Unlike the CNN model, these architectures did not undergo the same detailed optimization process involving fine-tuning and ablation studies. Instead, we applied the optimized hyperparameters outlined in Section 3.4.2 (Fine-tuned values column in Table 2) to these models. By applying the optimized hyperparameters from the CNN model to the training of other architectures, we maintained consistency across our benchmarking process, ensuring a fair comparison between models. This approach simplified the evaluation process while enhancing the reliability and reproducibility of our findings. Following this, we carried out the steps of model training,

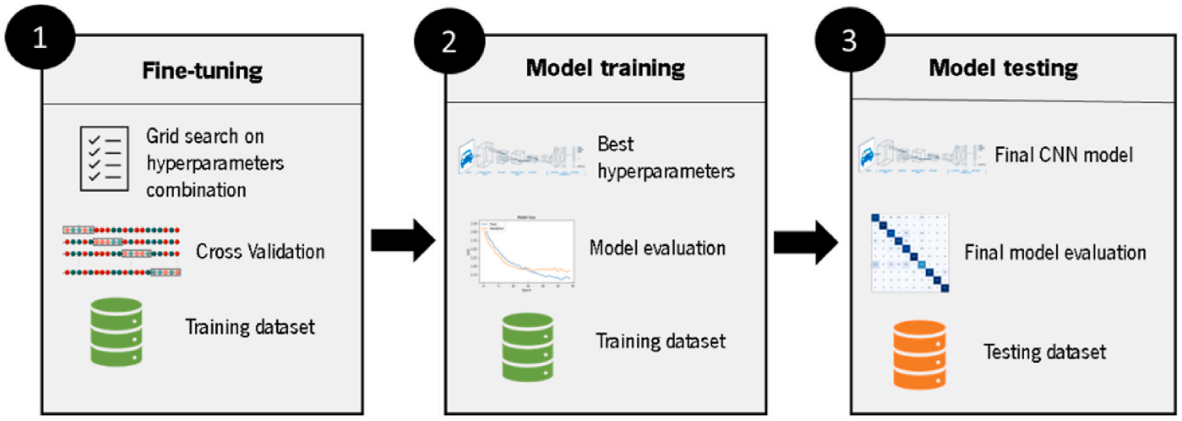


Fig. 6. Representation of the three stages for building the proposed CNN-based model.

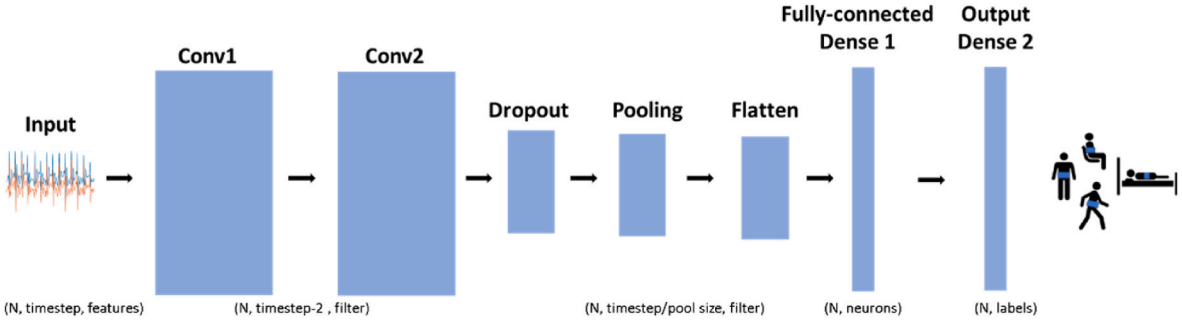


Fig. 7. Proposed CNN-based model architecture for daily motor activities recognition.

Table 3
Hyperparameters explored in the fine-tuning stage and best combination obtained for the final model.

Hyperparameters	Values	Fine-tuned values
Window size	5, 56, 112 and 240	240
Number of filters	32, 34 and 128	32
Pooling size	2	2
Number of neurons	60 and 100	60
Batch size	256 and 512	256
Dropout rate	0.3 and 0.5	0.5
Learning rare	0.001 and 0.0001	0.0001

validation, and testing as depicted in Fig. 6.

3.4.5.1. LSTM-based models. We developed two LSTM-based models, both founded on a similar structure consisting of two LSTM layers, as shown in Fig. 9. These layers are crucial for recognizing temporal sequences, which is essential for interpreting sequential patterns in data (Vrigkas et al., 2015), (Gupta et al., 2022), (Martins et al., 2022). The first model uses unidirectional LSTM layers, processing data in forward chronological order. In contrast, the second model employs bidirectional LSTM layers, analyzing data both forwards and backwards in time. This bidirectional approach improves the model's capacity to understand complex sequences by considering temporal relationships in both directions (Martins et al., 2022).

To prevent overfitting, we incorporated a dropout layer after each LSTM layer. Following these LSTM layers, we added a fully connected layer to introduce non-linearity, which is vital for the accurate classification of sequential data. The architecture concludes with another fully connected layer that acts as the classifier, converting the extracted features into predictions for the 6 defined classes.

3.4.5.2. Hybrid CNN-LSTM models. In our study, we introduced three CNN-LSTM architecture designs, depicted in Fig. 10, panels a) to c). The configuration of each architecture is outlined as follows. Initial Feature Extraction: At the forefront of each model, two 1D convolutional layers

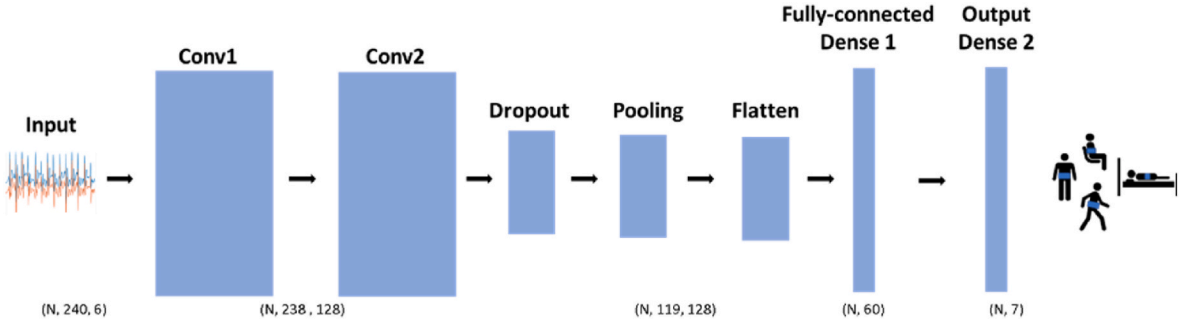


Fig. 8. Final CNN-based model architecture for daily motor activities recognition.

Table 4
Ablation study for the proposed final model architecture.

Model	Architecture
CNN-based model	C1 + C2 + D + MP + FL + FC1 + FC2
Ablation 1	C1 + D + MP + FL + FC1 + FC2
Ablation 2	C1 + D + MP + FL + FC2
Ablation 3	C1 + MP + FL + FC2
Ablation 4	C1 + FL + FC2
Ablation 5	C1 + BN + D + C2 + BN + D + MP + FL + FC1 + FC2

C: Convolution layer; D: Dropout layer; MP: Max pooling layer: FL: Flatten layer; and FC: Fully connected layer; BN: Batch Normalization Layer.

are employed alongside batch normalization and dropout layers. This setup is crucial for extracting spatial features from the input data. Spatial Information Consolidation: A global pooling layer is utilized to condense the spatial information extracted across the sequence, ensuring a streamlined representation for subsequent processing. Temporal Pattern Analysis: To capture the dynamic temporal dependencies within the sequential data, we integrate two BiLSTM blocks. This arrangement is pivotal for understanding the temporal aspects of the sequences. Feature Fusion Approaches: Each model incorporates one of three distinct feature fusion techniques, informed by prior CNN-LSTM research (Romijnders et al., 2023), (Kiprijanovska et al., 2020), (Altuve et al.,

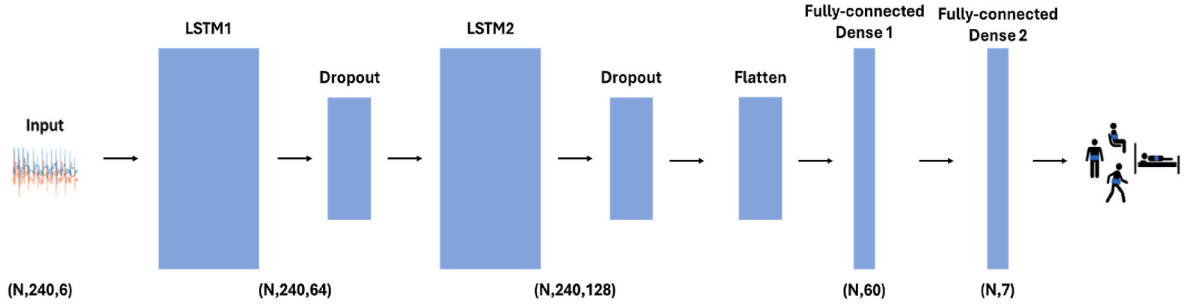


Fig. 9. LSTM-based general model, where LSTM blocks were considered uni and bidirectional.

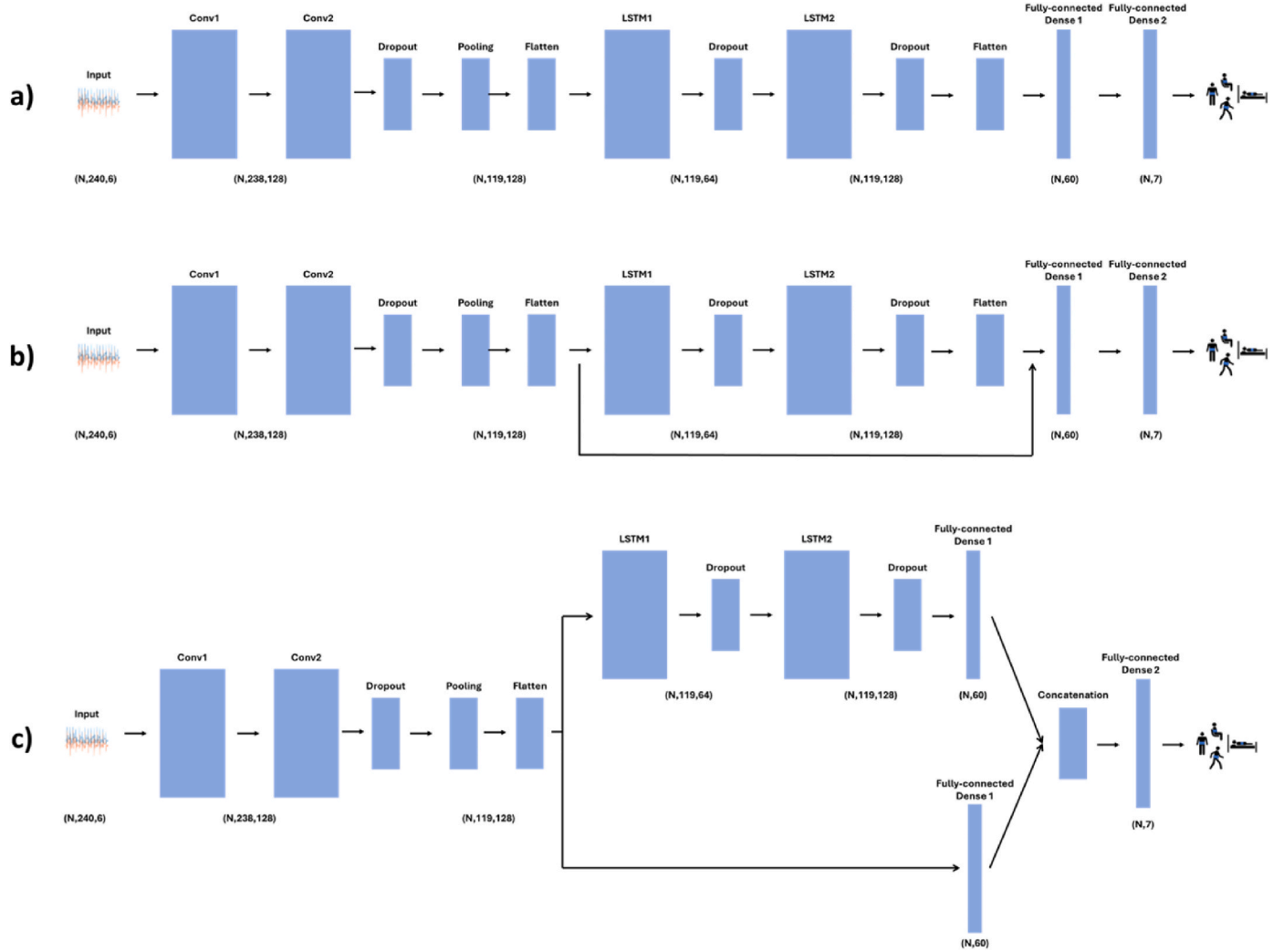


Fig. 10. Representations of the three proposed CNN-LSTM based models where a), b) and c) illustrate the distinct feature fusion strategies implemented.

2020). The strategies include: A straightforward approach that directly learns features without additional fusion mechanisms (Fig. 10 a)), A method that employs a residual block for the integration of CNN and LSTM-derived features (Fig. 10 b)), A technique that processes features from CNN and LSTM layers independently using fully connected layers prior to their concatenation (Fig. 10 c)). These fusion techniques are designed to optimally combine spatial and temporal features, aiming to improve the model's performance. Feature-to-Output Transformation: An intermediary fully connected layer facilitates the transition of the fused features into a suitable format for classification. Classification Output: The concluding output layer is responsible for translating the refined features into classifications across six predefined activity categories.

This structure aims to effectively harness both spatial and temporal data characteristics, enhancing the architectures' ability to accurately classify activities within the CNN-LSTM framework.

4. Results

In this section, we unveil the findings of our experiments. These were conducted offline using Python 3.9 on a Lenovo Legion Y540 system. The system specifications include an Intel Core i5, 9th Gen processor, NVIDIA® GeForce® GTX 1650 graphics card, 8 GB DDR4 memory operating at 2666 MHz, and a 512 GB SSD PCIe storage.

4.1. Training performance

Through the optimization of the algorithm performance by performing a grid search during the hyperparameters fine-tuning stage, it was found that the DL model obtained better performance using windows of 240 samples, organized into batches of 256 windows each, 128 filters in each convolutional layer, a pooling size of 2, 60 neurons in each dense layer, a dropout rate of 0.5 and the algorithm should update at 0.0001 rates. Then, the model was trained with fine-tuned parameters, and it was analyzed the F1 score and loss function curve during the training process, as depicted in Fig. 11. Table 5 shows the model performance during the training phase with 100 epochs.

4.2. Testing performance

To evaluate the performance of the proposed DL model, was used unknown data to the model (*testing dataset*). Table 6 shows the final model performance in the testing phase, and Fig. 12 depicts the obtained confusion matrix.

4.3. Ablation index performance

Table 7 displays the outcomes from the ablation study conducted on the CNN-based model architecture. The ablated models were trained and validated with the training and validation dataset, respectively, with the

Table 5
Performance of the final model during the training phase.

Epochs	Dataset	Loss value	Accuracy	Precision	Recall	F1 score
100	Training	0.278	0.912	0.927	0.899	0.913
	Validation	0.363	0.901	0.928	0.883	0.905

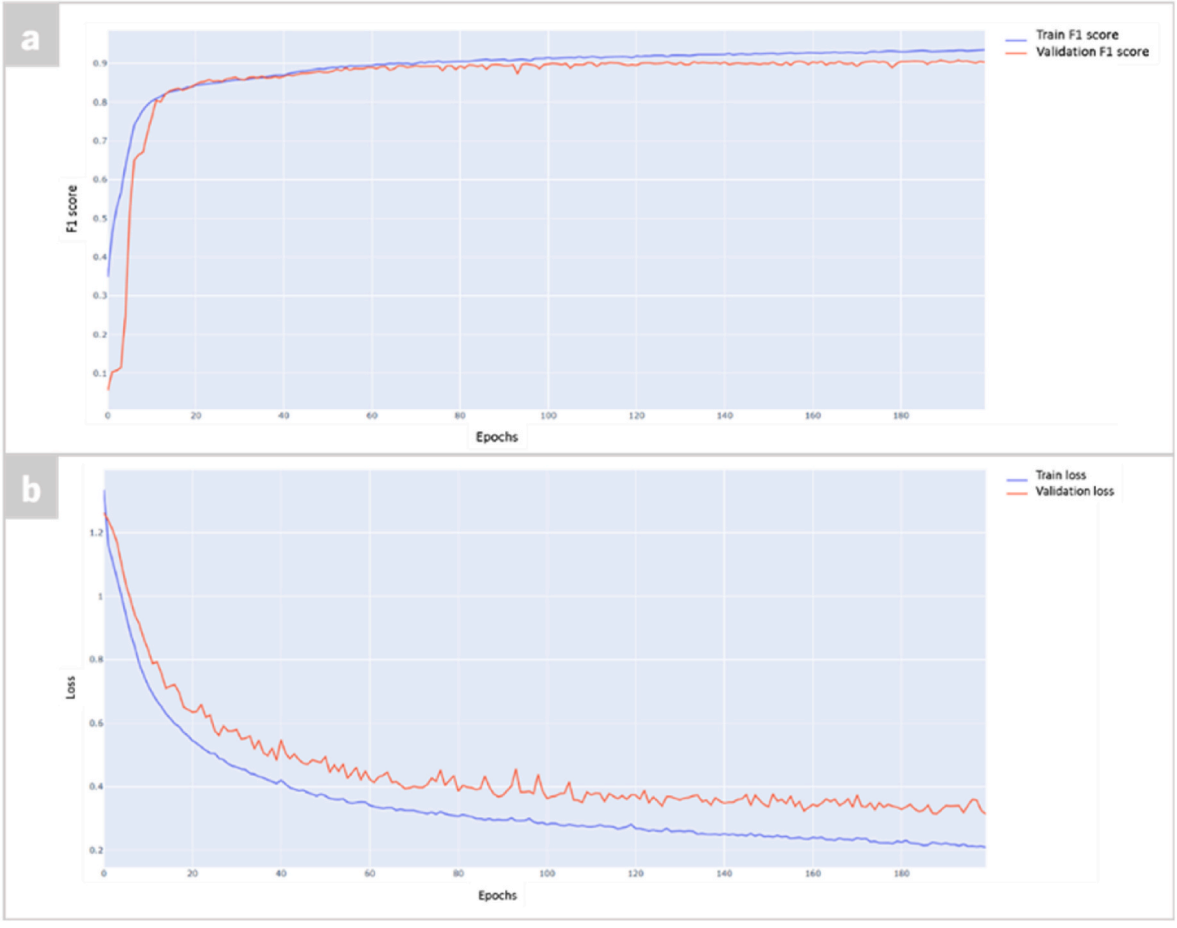


Fig. 11. Model performance during training: (a) F1 score value evolution for train and validation dataset; and (b) Loss function value evolution for *training* and *validation* dataset.

Table 6
Performance of the final model during the testing phase.

Dataset	Loss value	Accuracy	Precision	Recall	F1 score
Test	0.326	0.907	0.945	0.876	0.909

same parameters as the defined CNN-based model architecture. Test results were derived from the test dataset, and based on the performance, we selected the optimized CNN-based model for the subsequent benchmarking stage. Given the performance results of 91.1 %, 0.912, 0.911 and 0.906 for accuracy, precision, recall and F1 score, respectively, the model derived from Ablation 5 was selected as the optimized CNN-based model for the benchmarking stage.

4.4. Deep learning models benchmark results

To further evaluate the performance of our optimized CNN-based model, we carried out a benchmarking analysis against five models based on LSTM and CNN-LSTM designs. Table 8 displays the performance metrics obtained during the testing phase. Table 9 presents the computational costs associated with each model. This comprehensive comparison provides valuable information on the performance and computational efficiency of each architecture, supporting the selection process of the most suitable model.

5. Discussion

Although several investigations have been dedicated to developing AI-based computational methods for HAR, there is still no consensus on which model architecture is suitable (Rav et al., 2017)– (Zhang et al., 2022), (Murad and Pyun, 2017)– (Zainudin et al., 2016). However, the literature review on related works highlights that DL models can outperform classic ML, and around 25 % higher performance can be obtained with LSTMs and CNNs, as observed in (Zhang et al., 2022) and (Xu et al., 2019). Thus, a DL-based model architecture should be followed. This research explored, implemented and benchmarked a CNN-based model, along with other LSTM and hybrid CNN-LSTM based frameworks, capable of recognizing basic daily living activities performed by PD patients. Therefore, this research innovates by introducing a new HAR customized to PD, poorly addressed in HAR-related SoA (Tahavori et al., 2017).

During training, the CNN-based model, fine-tuned through grid search, demonstrated strong performance on the validation dataset, achieving an F1 score of 0.905, as shown in Table 5. We evaluated the model's performance across both the training and validation datasets, with the results for each training step visually depicted in Fig. 9. The learning curves suggested a well-fitted model, evidenced by lower loss

Table 8
Benchmark performance metrics test results.

Model	Dataset	Loss value	Accuracy	Precision	Recall	F1 score
LSTM	Test	0,388	0,885	0,879	0,885	0,874
BiLSTM		0,425	0,879	0,892	0,879	0,879
CNN-LSTM 1		0,314	0,892	0,886	0,892	0,883
CNN-LSTM 2		0,298	0,912	0,913	0,912	0,907
CNN-LSTM 3		0,303	0,914	0,913	0,914	0,908
Optimized CNN		**0,327**	**0,911**	**0,912**	**0,911**	**0,906**

Table 9
– Benchmark computational costs test results.

Architecture	Trainable Parameters	Memory (MB)	Train Time (s)	Test Time (s)
LSTM	332 K	1331	251,98	1,20
BiLSTM	763 K	3056	437,69	1,59
CNN-LSTM 1	450 K	1800	153,92	1,04
CNN-LSTM 2	450 K	1804	182,10	1,42
CNN-LSTM 3	473 K	1894	161,64	1,15
Optimized CNN	**52,9 K**	**0,211**	**127,43**	**0,73**

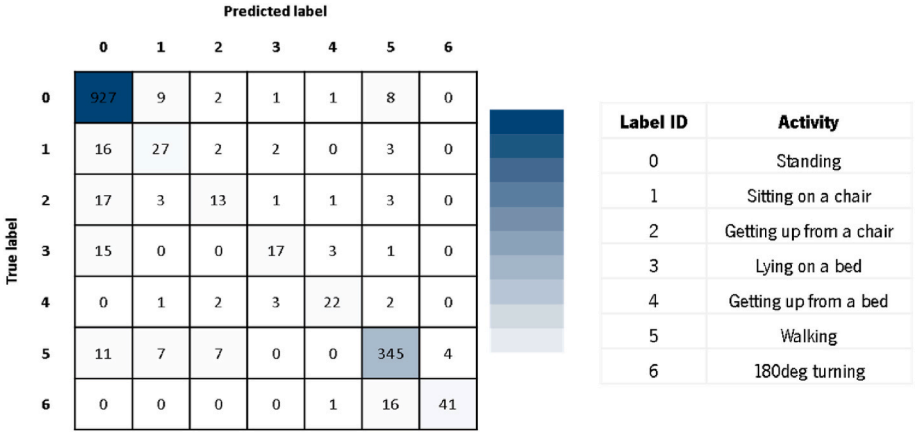


Fig. 12. Confusion matrix of the final model during the testing phase.

Table 7
Ablation study results.

		Model architectures					
		Final model	Ablation 1	Ablation 2	Ablation 3	Ablation 4	Ablation 5
Metrics	**Loss**	0.3262	0.488	0.492	0.405	0.361	0,327
	Accuracy	0.9074	0.898	0.8930	0.887	0.902	0,911
	Precision	0.946	0.949	0.939	0.912	0.924	0,912
	Recall	0.876	0.848	0.850	0.866	0.879	0,911
	F1 score	0.909	0.895	0.892	0.889	0.901	0,906

values on the training dataset than on the validation dataset, consistent with findings from prior studies (Goetz et al., 2008), (Brownlee, 2018), (Folador et al., 2019).

The CNN-based model was evaluated using the testing dataset not previously seen, where it recorded an F1 score of 0.909 and an accuracy of 90.7 %, as detailed in Table 6. When compared to similar studies utilizing CNN architectures, our accuracy was lower: 98.6 % in study (Rav et al., 2017), 98.2 % in (Zhang et al., 2022), and 92.55 % in (Cheng et al., 2023). Several factors contribute to this discrepancy of up to approximately 8 % in accuracy. First, the referenced studies included a wider range of daily activities in their datasets, such as jumping up and down stairs, sleeping, and using an elevator (Rav et al., 2017), (Zhang et al., 2022), (Cheng et al., 2023). Additionally, the study in (Rav et al., 2017) used an Inertial Measurement Unit (IMU) positioned at the hip, which may not accurately capture full gait cycles and could be more susceptible to movement artifacts. Placing the IMU on the lower trunk is preferred because it captures the complete gait cycle and is less influenced by user movements. The study in (Cheng et al., 2023) utilized 26 IMUs, which could reduce the system's acceptability and usability due to the increased demand for computational memory. Lastly, our results are derived from a dataset specifically curated for PD, contrasting with the studies (Goetz et al., 2008), (Brownlee, 2018), (Folador et al., 2019) that used data from healthy individuals.

When comparing the performance of the implemented CNN-based model with (Tahavori et al., 2017), non-significant differences were found in model accuracy (91.2 % in the proposed solution *vs* 92.99 % in (Tahavori et al., 2017)). Also, in (Tahavori et al., 2017), the dataset included data captured from 2 participants with PD, while 18 PD patients were included in this study. In addition, given that these datasets had unbalanced data, it could be more beneficial to use the F1 score for evaluating model quality instead of accuracy (Rav et al., 2017), (Zhang et al., 2022), (Rueda et al., 2018), (Tahavori et al., 2017).

The tasks of standing and walking were the activities most accurately recognized by the CNN-based model (see Fig. 10). These results were expected since these activities were most represented in the created dataset as depicted in Fig. 5. The continuous and cyclic nature of the data from these two activities allowed the model to learn well for recognizing these actions. It was observed that the CNN-based model confuses the activity of 180deg turning with walking because the inertial signals of these activities are very similar, which could be overcome using more features (see Fig. 10). The transitional activities of sitting on a chair, lying on a bed, and getting up from a chair/bed are shorter than dynamic tasks such as walking. Therefore, these labels were less represented in the dataset, making it more difficult for the model to learn their recognition (see Fig. 10). To overcome this constraint, a 50 % overlap between consecutive windows was used. Indeed, overall, the final model presented a positive performance. However, in the future, it could be beneficial to implement a hierarchical framework capable of firstly distinguishing dynamic (walking or turning) from static activities (sitting on a chair, lying on a bed, getting up from a chair/bed, or standing), and between each classification run another model to specify the motor activity.

The ablation study enabled to study the importance of each layer in the CNN-based model. The initial CNN-based model maintained the best results at each layer remotion, with an F1 score equal to 0.909 (Table 7). The decrement in F1 score, accuracy, precision, and recall was visible throughout the layer reduction process. The most ablated model obtained a still good F1 score of 0.901, and it could also be a possibility if a lighter model should be needed. During the last phase of ablation, we added Batch Normalization layers in conjunction with dropout layers to our model. This adjustment resulted in a slight increase in accuracy, achieving a value of 0.911. However, it coincided with a minor reduction in the F1 score, which decreased to 0.906. Since this model (Ablation 5 in Table 7) gave better results comparatively to the CNN-based model and the other ablations, from now how, it will be named of optimized CNN model.

In our analysis, we compared the performance of our optimized CNN model against other prevalent DL architectures using the same dataset, as detailed in Tables 8 and 9 This comparison revealed that our optimized CNN model surpassed both the LSTM-based models and the initial CNN-LSTM model (labeled as CNN-LSTM 1 in Table 8) in all evaluated performance metrics. Despite its overall strong performance, the optimized CNN model did perform slightly lower than the more advanced CNN-LSTM models. Specifically, it achieved an F1 score of 0.906, which is marginally less than the F1 scores of 0.907 and 0.908 recorded for the CNN-LSTM 2 and CNN-LSTM 3 models, respectively, as reported in Table 8.

Moreover, despite being slightly outperformed by more complex hybrid models, the optimized CNN model stands out due to its significantly lighter weight. According to Tables 9, it is approximately 9 times smaller and possesses 420K fewer trainable parameters than the top-performing model. Thus, considering the potential for further weight reduction through layer ablation (Table 7) while maintaining a significant portion of the model's performance, our optimized CNN model emerges as a highly viable option for HAR in PD patients when compared to the other DL frameworks.

5.1. Strengths and limitations

A CNN-based model for daily motor activities recognition in PD was implemented. The algorithm used inertial data capture from a single IMU placed in L5, using sensBand. The model showed satisfactory performance for recognizing walking, 180deg turning, standing, sitting on a chair, getting up from a chair, lying on a bed, and getting up from a bed. Furthermore, a custom dataset was created from data collection with eighteen patients with PD while performing trivial motor activities.

Even though positive results have been obtained, future challenges were identified considering the limitations of the proposed solution. Firstly, more data should be collected to enhance the learning capacity of the model. Also, it could be interesting to explore data-balancing techniques. Second, other model architectures can be explored, such as transformers (also suitable to lead with time-series classification), combining different hyperparameters, or even other window sizes and values of overlap. Moreover, adding other activities, such as loading bags, could be interesting to obtain a more comprehensive assessment of patients' motor performance. Future challenges also include investigating preprocessing methods, such as splitting accelerometer data into gravitational and body motion components, and exploring a hierarchical architecture based on two models—one for distinguishing dynamic from static activities and the second for specifying the motor activity.

6. Conclusions

This study aimed to implement and test a DL-based model for daily motor activities recognition based on inertial data to contribute for motion monitoring and assessment in PD. From here, it is expected to contribute to the PD field with an efficient system that provides continuous monitoring and assessment of patients' motor profile by recognizing trivial daily motor activities for further analysis of their performance and suitable therapeutic intervention prescription. For that, a custom dataset was created and prepared. Also, an end-to-end model based on a CNN architecture implementation was presented, which allowed to verify the formulated hypothesis following an extensive process of validation and benchmark against alternative DL frameworks. The implemented CNN-based model showed good performance for recognizing walking, standing, 180deg turning, sitting, getting up from a chair, lying, and getting up from a bed with a custom dataset. Future challenges should cover exploring other models and increasing the data population.

CRediT authorship contribution statement

Helena R. Gonçalves: Writing – original draft, Methodology, Conceptualization. **Luís Martins:** Writing – review & editing, Validation, Software, Methodology. **Nuno Ferrete Ribeiro:** Writing – review & editing, Validation, Software, Methodology. **Luís Abreu:** Writing – original draft, Validation, Software, Methodology, Conceptualization. **Ana Margarida Rodrigues:** Writing – review & editing, Writing – original draft. **Cristina P. Santos:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Conceptualization.

Funding

This work was supported by FCT national funds, under the national support to R&D units grant, through the reference project UIDB/04436/2020 and UIDP/04436/2020, and under the Reference Scholarship under grants SFRH/BD/136569/2018 and 2022.14549.BD.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Cristina P. Santos reports financial support was provided by University of Minho. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Altuve, M., Lizarazo, P., Villamizar, J., 2020. Human activity recognition using improved complete ensemble EMD with adaptive noise and long short-term memory neural networks. Biocybern. Biomed. Eng. 40 (3), 901–909. https://doi.org/10.1016/j.bbe.2020.04.007.

Baskar, S., Mohamed Shakeel, P., Kumar, R., Burhanuddin, M.A., Sampath, R., 2020. A dynamic and interoperable communication framework for controlling the operations of wearable sensors in smart healthcare applications. Comput. Commun. 149 (September 2019), 17–26. https://doi.org/10.1016/j.comcom.2019.10.004.

Batool, M., Jalal, A., Kim, K., 2019. Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm. Int. Conf. Appl. Eng. Math. ICAEM 2019 - Proc. 145–150. https://doi.org/10.1109/ICAEM.2019.8853770, 2019.

Brownlee, J., 2018. 1D convolutional neural network models for human activity recognition. Deep Learning for Time Series.

Cheng, D., Zhang, L., Bu, C., Wang, X., Wu, H., Song, A., 2023. ProtoHAR: prototype guided personalized federated learning for human activity recognition. IEEE J. Biomed. Heal. Inf. 27 (8), 3900–3911. https://doi.org/10.1109/JBHI.2023.3275438.

Dijkstra, B., Zijlstra, W., Scherder, E., Kamsma, Y., 2008. Detection of walking periods and number of steps in older adults and patients with Parkinson's disease: accuracy of a pedometer and an accelerometry-based method. Age Ageing 37 (4), 436–441. https://doi.org/10.1093/ageing/afn097.

Folador, J.P., Souza Santos, M.C., Oliveira Andrade, A., 2019. System Usability Scale (SUS) applied to a webbased integrated system for the management of data from people with Parkinson's disease. https://doi.org/10.5281/zenodo.3459740.

Garbin, C., Zhu, X., Marques, O., 2020. Dropout vs. batch normalization: an empirical study of their impact to deep learning. Multimed. Tool. Appl. 79 (19), 12777–12815. https://doi.org/10.1007/s11042-019-08453-9.

Garcia, K.D., et al., 2021. An ensemble of autonomous auto-encoders for human activity recognition. Neurocomputing 439, 271–280. https://doi.org/10.1016/j.neucom.2020.01.125.

Garcia-Gonzalez, D., Rivero, D., Fernandez-Blanco, E., Luaces, M.R., 2020. A public domain dataset for real-life human activity recognition using smartphone sensors. Sensors 20 (8). https://doi.org/10.3390/s20082200.

Gil-Martín, M., San-Segundo, R., Fernández-Martínez, F., Ferreiros-López, J., 2020. Improving physical activity recognition using a new deep learning architecture and post-processing techniques. Eng. Appl. Artif. Intell. 92, 103679. https://doi.org/10.1016/j.engappai.2020.103679.

Goetz, C.G., et al., 2008. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov. Disord. 23 (15), 2129–2170. https://doi.org/10.1002/mds.22340.

Gonçalves, H.R., Rodrigues, A., Santos, C.P., 2021. Gait monitoring system for patients with Parkinson's disease. Expert Syst. Appl. 185, 115653. https://doi.org/10.1016/j.eswa.2021.115653.

Gonçalves, H.R., Branquinho, A., Pinto, J., Rodrigues, A.M., Santos, C.P., 2024. Digital biomarkers of mobility and quality of life in Parkinson's disease based on a wearable motion analysis LAB. Comput. Methods Progr. Biomed. 244, 107967. https://doi.org/10.1016/j.cmpb.2023.107967.

Guan, Y., Plötz, T., 2017. Ensembles of deep LSTM learners for activity recognition using wearables. Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol 1 (2), 1–28. https://doi.org/10.1145/3090076.

Gupta, N., Gupta, S.K., Pathak, R.K., Jain, V., Rashidi, P., Suri, J.S., 2022. Human Activity Recognition in Artificial Intelligence Framework: a Narrative Review, vol. 55. Springer, Netherlands. https://doi.org/10.1007/s10462-021-10116-x, 6.

Huang, W., Zhang, L., Wu, H., Min, F., Song, A., 2023. Channel-Equalization-HAR: a light-weight convolutional neural network for wearable sensor based human activity recognition. IEEE Trans. Mobile Comput. 22 (9), 5064–5077. https://doi.org/10.1109/TMC.2022.3174816.

Inoue, M., Inoue, S., Nishida, T., 2018. Deep recurrent neural network for mobile human activity recognition with high throughput. Artif. Life Robot. 23 (2), 173–185. https://doi.org/10.1007/s10015-017-0422-x.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR abs/1502 (0) [Online]. Available: http://arxiv.org/abs/1502.03167.

K, M., Ramesh, A., G, R., Prem, S., A A, R., Gopinath, D.M.P., 2021. 1D Convolution approach to human activity recognition using sensor data and comparison with machine learning algorithms. Int. J. Cogn. Comput. Eng. 2, 130–143. https://doi.org/10.1016/j.ijcce.2021.09.001.

Kazemimoghadam, M., Fey, N.P., 2022. An activity recognition framework for continuous monitoring of non-steady-state locomotion of individuals with Parkinson's disease. Appl. Sci. 12 (9), 1–14. https://doi.org/10.3390/app12094682.

Kiprijanovska, I., Gjoreski, H., Gams, M., 2020. Detection of gait abnormalities for fall risk assessment using wrist-worn inertial sensors and deep learning. Sensors 20 (18). https://doi.org/10.3390/s20185373.

Kwon, Y., Kang, K., Bae, C., 2014. Unsupervised learning for human activity recognition using smartphone sensors. Expert Syst. Appl. 41 (14), 6067–6074. https://doi.org/10.1016/j.eswa.2014.04.037.

Martins, L.M., Ribeiro, N.F., Soares, F., Santos, C.P., 2022. Inertial data-based AI approaches for ADL and fall recognition. Sensors 22 (11). https://doi.org/10.3390/s22114028.

Moore, S.T., Dilda, V., Hakim, B., MacDougall, H.G., 2011. Validation of 24-hour ambulatory gait assessment in Parkinson's disease with simultaneous video observation. Biomed. Eng. Online 10 (September). https://doi.org/10.1186/1475-925X-10-82.

Murad, A., Pyun, J.Y., 2017. Deep recurrent neural networks for human activity recognition. Sensors (Switzerland) 17 (11). https://doi.org/10.3390/s17112556.

Rav, D., Wong, C., Lo, B., Yang, G., 2017. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE J. Biomed. Heal. Inf. 12 (1), 106–137.

Romijnders, R., et al., 2023. Ecological validity of a deep learning algorithm to detect gait events from real-life walking bouts in mobility-limiting diseases. Front. Neurol. 14. https://doi.org/10.3389/fneur.2023.1247532.

Rovini, E., Maremmani, C., Cavallo, F., 2017. How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review. Front. Neurosci. 11 (OCT). https://doi.org/10.3389/fnins.2017.00555.

Rueda, F.M., Grzeszick, R., Fink, G.A., Feldhorst, S., Ten Hompel, M., 2018. Convolutional neural networks for human activity recognition using body-worn sensors. Informatics 5 (2), 1–17. https://doi.org/10.3390/informatics5020026.

Tahavori, F., et al., 2017. Physical activity recognition of elderly people and people with Parkinson's (PwP) during standard mobility tests using wearable sensors. 2017 Int. Smart Cities Conf. ISC2 2017 8–11. https://doi.org/10.1109/ISC2.2017.8090858.

Vrigkas, M., Nikou, C., Kakadiaris, I.A., 2015. A review of human activity recognition methods. Front. Robot. AI 2 (NOV), 1–28. https://doi.org/10.3389/frobt.2015.00028.

Xu, C., Chai, D., He, J., Zhang, X., Duan, S., 2019. InnoHAR: a deep neural network for complex human activity recognition. IEEE Access 7, 9893–9902. https://doi.org/10.1109/ACCESS.2018.2890675.

Ye, H., De, S., 2017. Technology in Parkinson disease: challenges and opportunities. J. Int. Soc. Burn Inj. 43 (5), 909–932. https://doi.org/10.1002/mds.26642. Technology.

Yuan, H., et al., 2024. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. npj Digit. Med. 7 (1), 91. https://doi.org/10.1038/s41746-024-01062-3.

Zainudin, M.N.S., Sulaiman, M.N., Mustapha, N., Perumal, T., 2016. Activity recognition based on accelerometer sensor using combinational classifiers. In: ICOS 2015 - 2015 IEEE Conf. Open Syst., pp. 68–73. https://doi.org/10.1109/ICOS.2015.7377280.

Zanardi, A.P.J., et al., 2021. Gait parameters of Parkinson's disease compared with healthy controls: a systematic review and meta-analysis. Sci. Rep. 11 (1), 1–13. https://doi.org/10.1038/s41598-020-80768-2.

Zhang, S., et al., 2022. Deep learning in human activity recognition withWearable sensors: a review on advances. Sensors 22 (4). https://doi.org/10.3390/s22041476.

Zwartjes, D.G.M., Heida, T., Van Vugt, J.P.P., Geelen, J.A.G., Veltink, P.H., 2010. Ambulatory monitoring of activities and motor symptoms in Parkinsons disease. IEEE Trans. Biomed. Eng. 57 (11), 2778–2786. https://doi.org/10.1109/TBME.2010.2049573.