Biomedical Signal Processing and Control 75 (2022) 103617

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Classification of healthy, Alzheimer and Parkinson populations with a multi-branch neural network

José Francisco Pedrero-Sánchez [a], Juan-Manuel Belda-Lois [a,b], Pilar Serra-Añó [c,*], Marta Inglés [c,d], Juan López-Pascual [a]

[a] Instituto de Biomecánica de Valencia. Universitat Politècnica de València, edificio 9C. Camino de Vera, s/n, 46022 Valencia, Spain
[b] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Valencia, Spain
[c] UBIC, Departament de Fisioteràpia, Universitat de València, 46010 Valencia, Spain
[d] Freshage Research Group, Department of Physiotherapy, Universitat de València, Centro de Investigación Biomédica en Red Fragilidad y Envejecimiento Saludable (CIBERFES-ISCIII), Fundación Investigación del Hospital Clínico Universitario de Valencia (INCLIVA), 46010 València, Spain

ARTICLE INFO

Keywords:
Alzheimer disease
Parkinson's disease
Inertial sensor
Multi-branch convolutional classifier
Functional assessment

ABSTRACT

Signal processing, for delimitation of the target events and parametrization, is usually required when instrumented assessment is conducted to determine an individual's functional status. However, these procedures may rule out relevant information obtained by sensors. To prevent this, the use of models based on neural networks that automatically extract relevant features from the raw signal may improve the characterization of the functional status. Thus, the aim of the study was to determine the classification accuracy of a multi-head convolutional layered neural network (CNN) using a simple functional mobility test in people with different conditions. The raw data from an inertial sensor embedded in a smartphone worn by 90 volunteers (i.e. 30 volunteers with Alzheimer's disease, 30 with Parkinson's disease and 30 healthy elderly people) was obtained. The CNN classification accuracy was compared to that of the two parametric classifiers, namely, linear discriminant analysis and multilayer perceptron, a neural network-based classifier.

As a result, the validation process revealed that the CNN classifier correctly assigned 100% of the participants to each group. The best accuracy in pathology classification for the two parametric classifiers ranged from 55% to 88%.

Therefore, the CNN model provided enhanced classification accuracy as compared to the parametric approaches, even better than the neural network-based classifier. Non parametrization may increase relevant information, thus enhancing pathology impact characterization.

1. Introduction

Funtional testing is essential in clinical practice due to its efficacy for quantifying the functional status of patients and improving clinical decision making. This type of assessment evaluates the patient's ability to perform several tasks or daily life activities [1], such as walking [2], maintaining balance [3,4], sitting and getting up from a chair [5,6] or performing combinations of tasks [7,8].

Eventually, the clinical purposes of functional assessment procedures are to establish severity classifications of the disease [9] and determine functional status categories [10] or levels of risk [11,12]. These classifications allow a better characterization of the patients and therefore a more accurate planning of the therapeutic interventions, as well as better prevention strategy implementation, when possible. To this end, instrumented assessments using sensors have been recently developed, to automate the recording and to obtain quantitative data that provide more detailed information about task performance [13–17]. In this context, our research group has demonstrated the utility of inertial systems embedded in Android devices for the determination of the functional status of people with neurological diseases, such as Parkinson's or Alzheimer's [18,19].

There are many studies that have used the output of instrumented assessment to conduct clinical classifications using statistical methods such as discriminant analysis (LDA) [20], logistic regression [21], random forest [22], extreme learning machines (ELM) [23] or support vector machines (SVM) [24–26].

* Corresponding author.
E-mail address: pilar.serra@uv.es (P. Serra-Añó).

https://doi.org/10.1016/j.bspc.2022.103617
Received 16 November 2021; Received in revised form 10 February 2022; Accepted 27 February 2022
Available online 11 March 2022

These classification methods require three main steps: feature extraction from the sensor data, selection of the clinically relevant features and implementation of the classification algorithm [27].

Several drawbacks have been reported in feature extraction and selection. Features are based on expert knowledge on the topic and relevant information may be missed when addressing new problems [28]. Further, the accuracy of the resulting classification can also be conditioned by the selection mechanism of the set of clinically relevant variables [22]. Moreover, there may be significant differences in the results depending on the algorithms used in feature computation [29]. To prevent this, getting as many characteristics as possible from the sensor data and then selecting a subset of variables is the norm [30]. However, feature selection methods are often subjective and selected parameters can be highly correlated [27].

To avoid this, the use of artificial neural networks (ANN) may provide a solution due to their ability to solve highly complex problems with as little signal manipulation as possible. ANNs can be designed to compute almost every type of input data [31] with the particularity that the technique itself extracts the characteristics of interest from a training process [28,32].

Some authors have already shown the potential of using sensor raw data as ANN input for clinical applications [12,16,33]. However, only one study has compared the accuracy of two classification methods, one convolutional layered neural network (CNN) fed with sensor data, and the other, a feature-based method [16]. In this study, an accuracy greater than 90% in classifying the OFF and ON states in Parkinson s disease was obtained. As stated by the authors, these promising results encourage research on the design and evaluation of more elaborate CNN architectures with the aim of classifying people with different pathologies and conditions, further using embedded sensors in smart watches or phones.

To determine the usefulness for clinical classification of a sensor data-driven CNN, which does not require feature extraction, its accuracy should be compared with other ANN methods which do require parameter extraction, such as the multilayer perceptron (MLP). Further, comparison with other linear parametric methods, such as LDA, the simplest statistical parametric classifier, should also be recommended to compare the accuracy of an ANN-based method with that of a linear parametric design.

The main goal of this study was to determine the classification accuracy of a multi-modal convolutional layered neural network (CNN) designed to process the raw sensor data obtained from a functional mobility test. To this end, a sample composed of people with Alzheimer's disease, Parkinson's disease and healthy elderly people was used. As a secondary goal, the accuracy of this technique was compared with that obtained by statistical techniques that require prior definition and selection of computed parameters such as LDA and a simple version of a Multilayer perceptron (MLP) classifier.

2. Material and methods

2.1. Participants

The sample of this study included three groups of participants, previously recruited for other studies conducted by our group [18,19]. The inclusion and exclusion criteria were:.

- Thirty people diagnosed with Parkinson's disease (PD) according to the United Kingdom Bank Criteria [34]. The inclusion criteria were: diagnosis of mild to moderate PD (Hoehn & Yahr stage 2 and 3 [35]); optimized and stable pharmacological therapy for at least 1 month before enrolment and no cognitive impairment, as assessed by the Folstein Mini Mental State Exam (score $\geq$ 24) [36]. All participants were assessed during "on" state, one hour after taking medication.
- Thirty people diagnosed with Alzheimer (AD) using the revised NINCDS-ADRDA criteria [37]. The inclusion criterion was: classified as stages 1 and 2, according to the Clinical Dementia Rating (CDR) [38].
- Thirty age-matched controls without any associated mental or musculoskeletal pathology.

Volunteers of the three groups were excluded if they presented other neurological or orthopedic impairments limiting independent walking and sitting down or getting up from a chair and severe uncorrected visual or auditory disorders. People in the PD group having received Deep Brain Stimulation or Duodopa treatment were excluded.

All participants accepted and signed an informed consent form for enrollment in the study. For the sample of people with Alzheimer, written informed consent was obtained from their caregivers. The protocol fulfilled the principles of the World Medical Association's Declaration of Helsinki. Since this work is based on two previous studies [18,19], all the procedures included were previously approved by the **Ethical Committee of the University of Valencia (Nr. 1517239006520).**

2.2. Data acquisition

The functional test was performed using the Fallskip® system (Biomechanical Institute of Valencia, Valencia, Spain). Fallskip® is an Android-based recording system running on a smartphone. The system was attached to the participant's lower back at L4-L5 level by fastening a Velcro strap. The system records the signals from the embedded Inertial Measurement Unit (IMU) (High Performance 6-Axis MEMS MotionTracking^{TM} composed of 3-axis gyroscope; 3-axis accelerometer) at 100 Hz.

The assessment required the measurement and registering of the weight, height, gender and previous year's history of falls of each participant. These variables were used for subsequent analyses.

The measurement protocol has been previously validated [18,19,39] and consisted in a mobility test with 5 phases performed sequentially in a single recording (Supplementary material 1).

- **Phase 1.** Standing still, the arms alongside the body for 30 s.
- **Phase 2.** Walking straight ahead as fast and as safely as possible towards a chair 3 m away.
- **Phase 3.** Turning around, sitting down in a chair.
- **Phase 4.** Standing up from the chair.
- **Phase 5.** Walking back as fast and as safely as possible to the starting point.

2.3. Data processing

All the data processing was performed by means of custom specific offline Python (3.x) scripts. The sensor raw data were processed according to Nishiguchi *et al.* [40] and Zijlstra *et al.* [41]. First, we performed a linear interpolation to get a constant sampling rate for all signals. A low-pass Butterworth filter (fourth-order zero-lag at 20 Hz) was then applied. The data recorded by the smartphone sensors were processed to generate the inputs to all the classification models. For the subsequent analysis, different data processing was performed for the parametric methods and for the CNN as illustrated in Fig. 1.

2.3.1. Data processing for parametric classification Models, LDA and MLP

Functional test segmentation and parametrization was performed following the procedures described in Serra-Añó *et al.* [18,19].

2.3.1.1. Functional test segmentation. Each functional test recording was manually split up into the 5 phases described above by the same researcher. The beginning and the end of each phase was selected by clicking on a graph where accelerometers, gyroscopes and magnetometer data vectors were plotted.

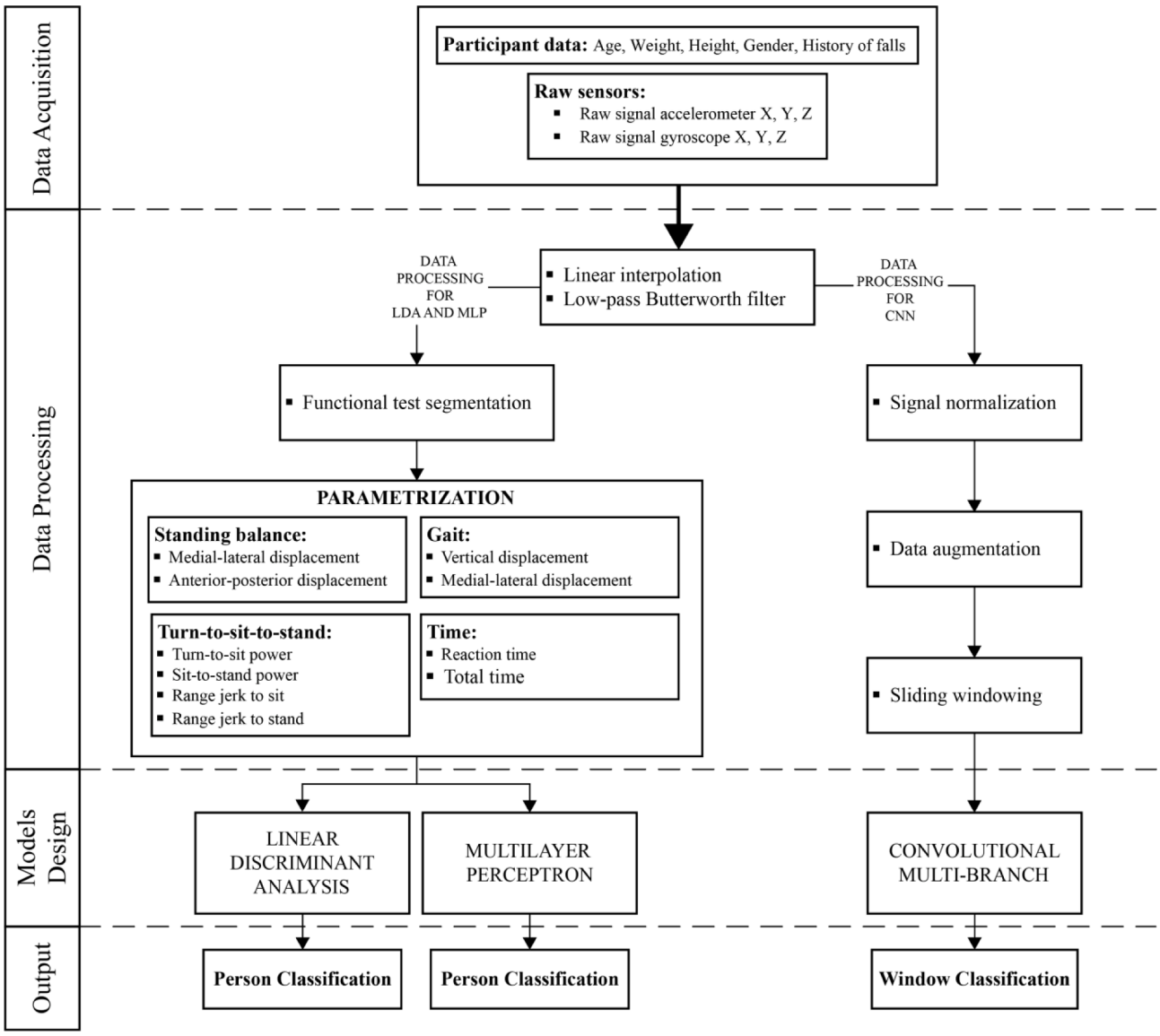

Fig. 1. Data processing flow chart for parametric (left) and non-parametric (right) classification models from the acquisition data to classification result. Parametric classification models directly classify persons and non-parametric classification models classify raw signal windows (not persons).

2.3.1.2. Parametrization. The variables calculated per phase and for the whole functional test have been previously validated [40–43] and were the following:.

- **Phase 1, balance:** range of the *Medial-lateral displacement (MLDisp)* of the center of mass (COM); range of *Anterior-posterior displacement (APDisp)* of the COM.
- **Phase 2 & 5, gait:** range of the *Vertical displacement (Vrange)* of the COM; range of the *Medial-lateral displacement (MLrange)* of the COM.
- **Phase 3 & 4, turn-to-sit-to-stand:** *Turn-to-sit power (PTurnSit)*; *Sit-to-stand power (PStand)* [44]; range of *jerk to sit (JerkSit)*; range of *jerk to stand (JerkStand)* [45].
- **Complete assessment:** Reaction time (*Reaction Time); Total time (Total Time)*.

2.3.2. Data processing for CNN

2.3.2.1. Signal normalization. Three signals were obtained from the accelerometer and three from the gyroscope, one for each axis (X, Y, Z). Each signal was normalized by the maximum range obtained in each sensor to provide homogeneous data inputs between 0 and 1 Eq. (1).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

2.3.2.2. Data augmentation. In order to increase the number of observations for CNN model training, artificial data was added to the pool through data augmentation techniques [46]. Data augmentation consisted of 90 and 180-degree signal rotation. Therefore, the total data was increased from 4 k observations per group to 12 k.

2.3.2.3. Sliding window. After the normalization of the signals and data augmentation, the data was reshaped as a hypermatrix. The hypermatrix consisted of windows having 64 samples of the six signals from the IMU with a 50% overlap following the window size and overlap recommendations for human activities according to Banos *et al.* [47] and Dehghani *et al.* [48]. Therefore, the size of the hypermatrix was Nx64x6, N being the number of windows.

2.4. Design of the classification models

The three classification models were designed to deal with two types

of input data as group predictors: i. Participants' data; and ii. Functional test data. The combination of these two types of data may contribute to increase the sensitivity of the classification, since some participant data (e.g., height, weight) are related with the magnitude of the biomechanical measures (e.g., balance, stand and sit) [41,44,49]. With the aim of avoiding bias in the classification models, a 1-way ANOVA with the between-subjects factor 'group' was used to search for differences in the participants data among groups. The type I error was set at 5% ($p \leq 0.05$). Variables significantly different among groups were discarded for model design.

All the statistical analyses were performed using RStudio software with R version 4.0.3. Keras API and Tensor Flow 2.0 in Python 3.7.x were used for MLP and CNN development.

2.4.1. Parametric classification models

Two parametric classification models were implemented, a lineal discriminant analysis (LDA) and a multilayer perceptron (MLP).

2.4.1.1. LDA classification model. In the LDA model, the participant's group was selected as the dependent variable while the participant data and the test parameters were introduced as independent variables. LDA was computed with Bayed approximation using the MASS package [50] with "*moment*" method for standard estimators of the mean and variance.

2.4.1.2. MLP classification model. A basic sequential model was designed using the participant's data and the test parameters as predictors and the group as the model output. The design consisted of an input layer of 12 neurons (one for each parameter), one dense hidden layer of 50 neurons with ReLU activation function and an output layer of three neurons. The output layer used a softmax function for the activation, transforming the sum of the predicted values from all the neurons in the output layer to 1. The compiler was defined as a categorical cross-entropy loss measure with Adam optimizer for a better efficiency. Finally, the evaluation metric accuracy was specified and an iterative design process was run, obtaining the best results for a configuration with a batch size of 32 for 200 training epochs.

2.4.2. Non-parametric classification models

The non-parametric model developed was a classification model based on convolutional neural networks (CNN) which used the participants' data and the raw data collected form the IMUs (i.e., 3D accelerometer and 3D gyroscope) as input data. The model was designed as two independent input data flow branches which were later concatenated to compute the group classification. The structure of the CNN model is shown in Fig. 2.

The participant data branch structure was designed as two dense layers with 500 and 100 neurons respectively followed by a transformer layer to perform the concatenation of two input data branches. The raw sensor signal branch structure was a two 1D convolutional layers with ReLU activation function, 64 filters each one and a kernel size of 3; followed by two layers that slow down the learning (dropout = 0.5 and max pooling = 2 layer). Finally, a transformer layer to perform the data concatenation (flatten layer) was used. For the top model, which automatically groups the features extracted from each branch (both flatten layers output), we used a structure of an MLP classification model with a dense layer with 100 neurons and an output layer with softmax activation function. To compile the model, we used the categorical cross-entropy, as our loss measure, and the Adam optimizer. Ultimately, we specified the evaluation metric "accuracy" and, to fit the model, an iterative design process was performed, obtaining the best results for a configuration with a batch size of 32 for 500 training epochs.

2.5. Data analysis

Classical statistical methods were used to obtain the mean as a central measure of trend and the standard deviation (SD) as a measure of dispersion. A 1-way ANOVA with the between-subjects factor 'group' was used to search for differences in the functional test variables. The type I error was set at 5% ($p \leq 0.05$).

Participants were randomly allocated to a training set, comprising 70% of participants and a validation set with the remaining 30% of participants (Fig. 3) to test the accuracy of the classification models and to compare them.

2.5.1. Development of the classification models

The classification models were designed with the training set. The calculation of the LDA resulted in two discriminant functions. The epochs were determined iteratively and the MLP model was obtained after training for 200 epochs, while the CNN was trained for 500 epochs. In both cases, the accuracy and loss evolution plots over the training epochs were obtained.

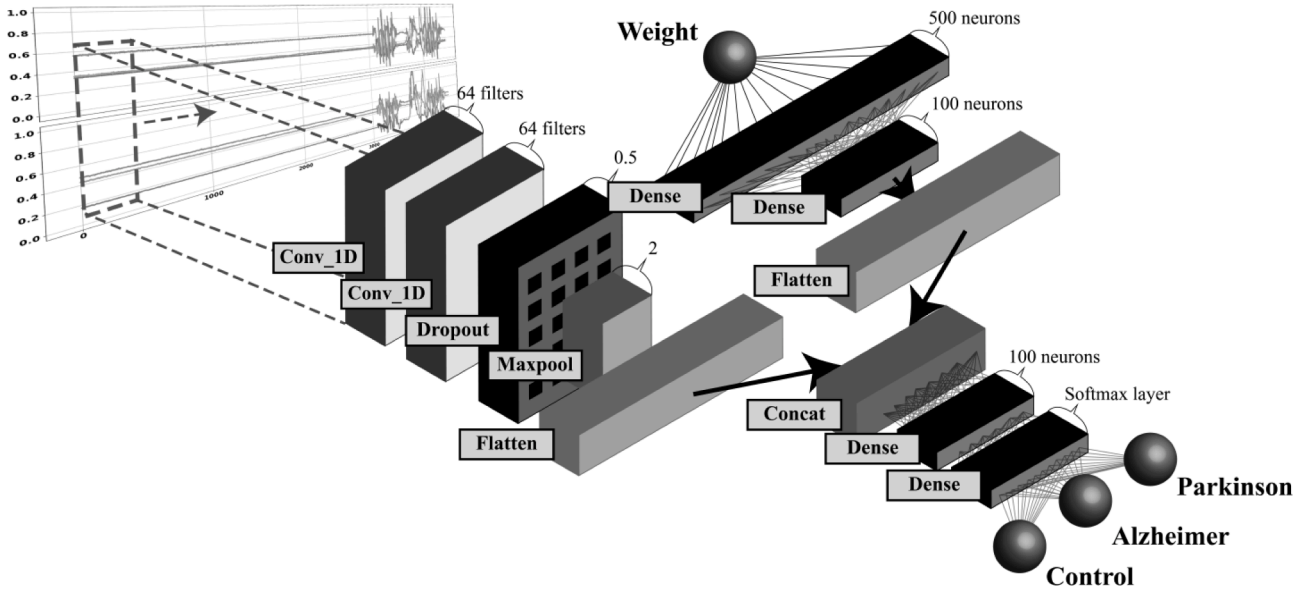

Fig. 2. Structure of convolutional neuronal network (CNN) multi-branch classification model. Left branch: input 64 samples sliding window of sensor (accelerometer XYZ and gyroscope XYZ). Right branch: input participant data.

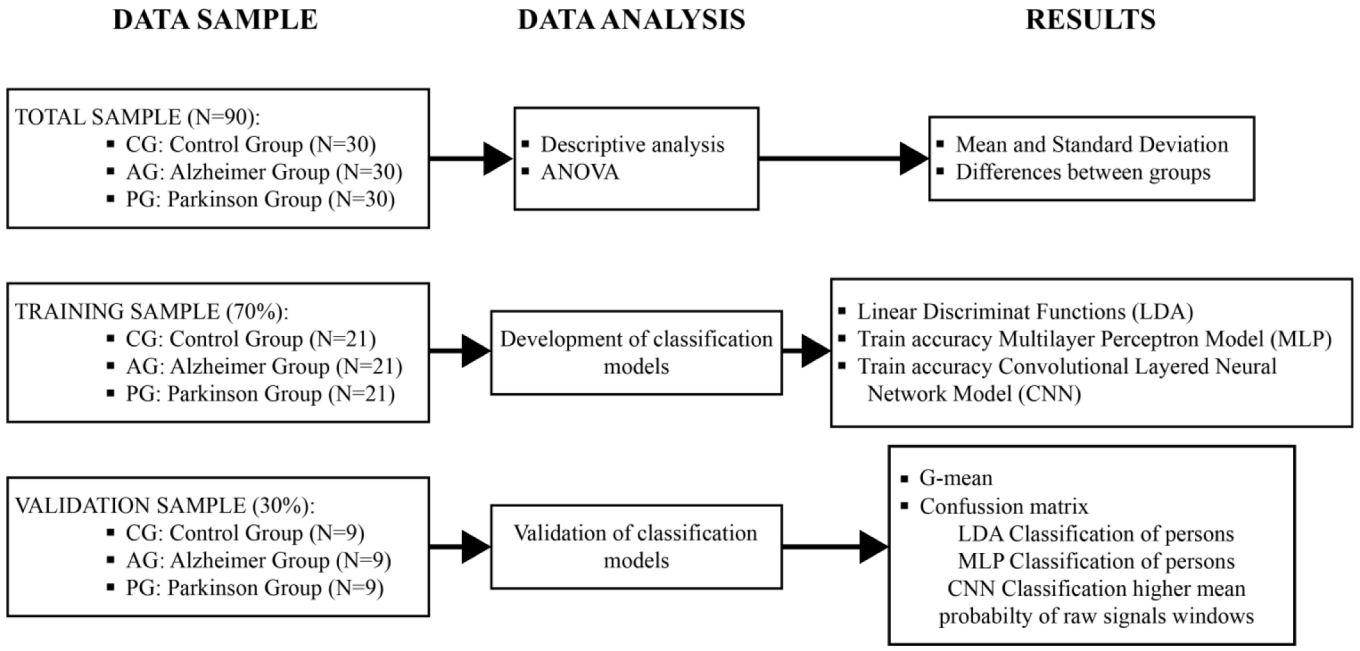

Fig. 3. Flow chart of data analysis with the mobility tests. Division of the total sample (90 participants) into two datasets: training set (70%) and validation set (30%). These sets were used to design and validate all classifiers: linear discriminant (LDA), multilayer perceptron (MLP) and convolutional layered neural network (CNN).

2.5.2. Validation of the classification models

The validation was conducted with the validation set. The accuracy of each classification model was assessed using a confusion matrix and the geometric mean (G-mean), that provides a robust metric of the quality of the classifier [51].

For the LDA and the MLP, the classifier provides a single classification per user, while the CNN provides a classification per window. The latter obtained an Nx3 array per participant with the probability of belonging to each of the 3 groups in the N windows in which the input data was split up. To get the group classification, the mean probability of each group across the N windows was computed, getting a 1x3 array per participant. Each participant was assigned to the group whose resulting probability was higher in the averaged windows. The confusion matrix for the CNN was obtained with this latter classification.

The comparison of the LDA and the MLP allowed to establish the differences in accuracy classification between a simple linear method and an ANN-based method.

The comparison of the two ANN-based methods (MLP and CNN) allowed to explore the potential differences of using a selection of parameters or the sensor raw data as the classification model input.

3. Results

3.1. Participants

Significant differences were found in the mean (SD) age of the participants (AG: 77.0 (8.63) years, PG: 67.3 (9.01) years and CG: 73.1 (11.80) years). There were also significant differences in height (AG: 1.55 (0.08) m, PG: 1.69 (0.07) m and CG: 1.71 (0.07) m). Further, significant differences in the sex distribution among groups were found ($\chi^2 = 14.28$; $p < 0.05$). Weight did not significantly differ ($p > 0.05$) between groups (AG: 72.4 (11.71) kg, PG: 75.7 (12.65) kg and CG: 72.5 (14.81) kg). Age, height and sex were not included as data input in the development of the classification models.

3.2. Descriptive analysis

Table 1 shows the descriptive values of the biomechanical parameters obtained in the assessment. As shown, there were significant differences between the groups in all variables except *MLrange* and *JerkStand*.

Table 1
Comparison of biomechanical parameters between control, Alzheimer and Parkinson's disease groups.

	CG ($n = 30$) Mean (SD)	AG ($n = 30$) Mean (SD)	PG ($n = 30$) Mean (SD)	ANOVA F value	*p*-value
MLDisp (mm)	3.76 (1.55)	7.86 (3.95)	13.82 (10.60)	17.7	**<0.001***
APDisp (mm)	14.14 (3.82)	23.96 (11.00)	28.37 (13.06)	15.6	**<0.001***
Vrange (mm)	30.45 (8.39)	20.15 (6.27)	27.96 (9.95)	12.4	**<0.001***
MLrange (mm)	56.72 (14.19)	63.54 (16.83)	56.72 (22.10)	1.4	0.244
PTurnSit (W)	132.99 (25.24)	52.00 (21.55)	77.11 (28.46)	80.9	**<0.001***
PStand (W)	315.74 (66.10)	146.05 (54.65)	214.13 (88.66)	43.1	**<0.001***
JerkSit (m/s^3)	21.88 (7.16)	16.28 (6.37)	19.84 (10.13)	3.7	**0.028***
JerkStand (m/s^3)	25.25 (8.21)	20.96 (6.87)	27.32 (15.99)	2.5	0.088
Reaction Time (s)	0.95 (0.42)	1.79 (0.82)	1.29 (0.38)	16.0	**<0.001***
Total Time (s)	11.78 (2.09)	20.56 (4.78)	15.30 (3.91)	41.4	**<0.001***

CG: control group; AG: Alzheimer group; PG: Parkinson Group; Data are expressed as mean (SD), * significant differences between groups ($p < 0.05$).

3.3. Development of the classification models

The coefficients of the variables included in the two discriminant functions resulting from the LDA analysis are presented in Supplementary material 2.

The accuracy evolution curves during the training of the MLP and CNN classification models are shown in Fig. 4. It is noted that the MLP

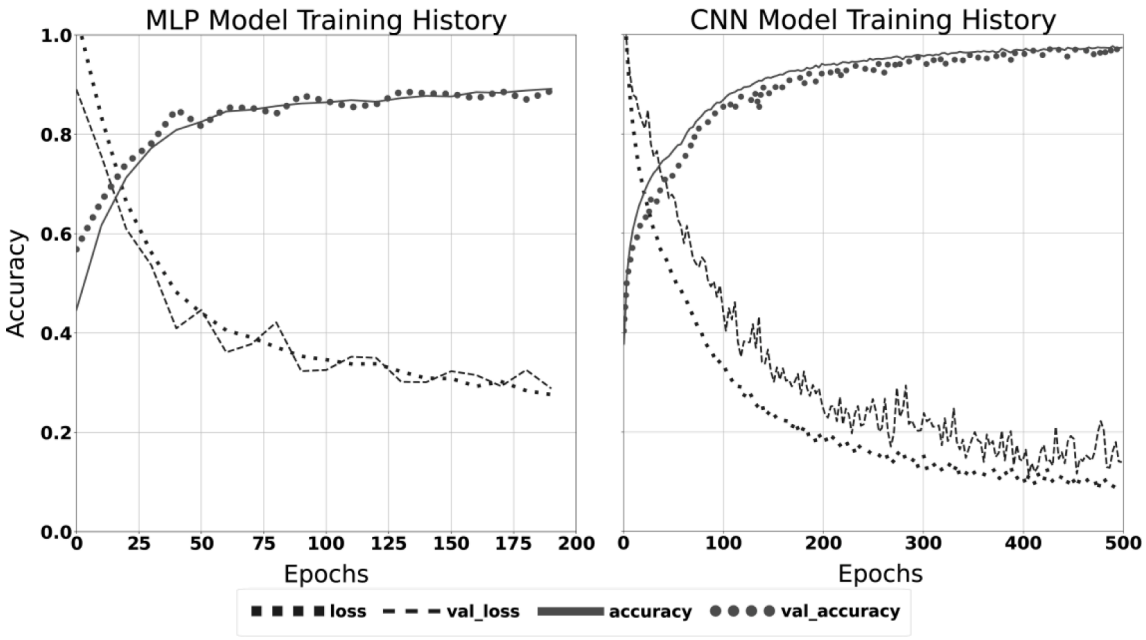

Fig. 4. Multilayer perceptron (MLP) (left) and Convolutional Neural Network (CNN) (right) training accuracy history models using the 63 training participants (70% of the total sample; 21 participants in each group). For a better visualization of the plot, the CNN training results were sub-sampled (1:3).

accuracy stabilizes at around 85%, after 150 epochs. In the CNN classifier, an accuracy higher than 95% is already reached at epoch 250.

3.4. Validation of the classification models

The accuracy of the classification models was evaluated with a validation sample of 27 participants (i.e., 9 in each group). Comparison of the accuracy of three models (i.e., LDA, MLP and CNN) is shown in Fig. 5.

LDA and MLP obtained almost the same accuracy results for the group of Alzheimer participants; both techniques correctly classified 77% and 80% of the participants, respectively. When Parkinson volunteers were analyzed, 55% of the participants were correctly classified using LDA and 80% with MLP. The control participants were correctly classified 100% using LDA and 88% with MLP. In contrast, the CNN model performed an accurate classification of all the 27 participants of the validation sample. The G-mean obtained are 0.83, 0.91 and 1.00 for the LDA, MLP and CNN classifiers respectively.

4. Discussion

This study presents the results for the classification accuracy of two neurological pathologies and their age-matched healthy counterparts with an analysis technique based on the use of convolutional neural networks (CNN), when performing a functional mobility test, in comparison with other traditional classification techniques.

It is noted that the CNN-based classification technique used allows classifying 100% of the subjects in Alzheimer's disease groups, Parkinson's disease groups and healthy controls, without also requiring the planning of any functional signal landmark selection or parametrization process. The designed technique outperforms the classification accuracy of MLP and LDA.

There are previous studies in which the assessment of functional mobility capacity has been used to classify pathologies or conditions

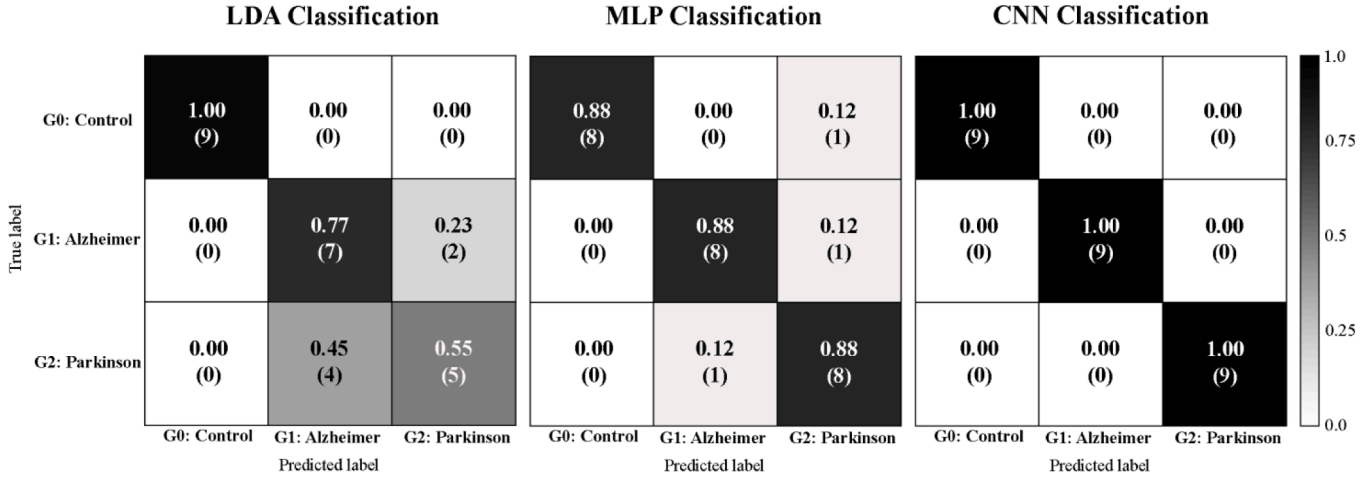

Fig. 5. Confusion matrices of the Linear Discriminant Analysis (LDA) (left), Multilayer perceptron (MLP) (center) and Convolutional Neural Network (CNN) (right) models using the 27 validation participants (30% of the total sample; 9 participants in each group). In parentheses, the number of observations classified in each group.

based on its efficacy in terms of planning the therapeutic approach. Indeed, Fritz *et al.* [1] compared motor performance in people with dementia, Parkinson's and Alzheimer's disease. An extensive protocol was followed including instrumented gait analysis (GAITRite), balance and dual-task assessment (Tinetti Mobility Test, Berg Balance Scale, Timed Up and Go and figure of 8 walk test) and hand dexterity (9 Hole Peg Test). In contrast to our results, they neither found differences between the AG and the PG in gait, nor in balance and dual-task assessment, but they did however report differences in hand dexterity. Since their sample size was similar, the fact that they failed to correctly classify the included pathologies was probably due to them mainly using non-sensored measurements.

This may be supported by the results of Satyabrata Aich *et al.* [22] who successfully classified people in the AG and the PG using spatio-temporal gait data collected through wearable devices and 3D Motion Analysis System. In this study, a comparison between several parametric classifiers (i.e., LDA NaiveBayes, Random Forest and Support Vector Machine (SVM)) was conducted. Due to the complexity of managing a high number of parameters, the authors applied non-lineal reduction algorithms. The best combination of parameters achieved 92.59% accuracy using the SVM classification. The higher accuracy of this classification model compared to the two parametric approaches presented in our manuscript may be related to the more sophisticated lab assessment performed, including 3D kinematics.

In this regard, some studies concluded that parametric analyses require not only a high number of parameters to define functional status, but also complex user instrumentation [17], several body sensors being needed. This combination of increased number of sensors and parameters has been claimed as necessary so as not miss relevant information on the functional status, which may jeopardize the power of classification for different stages of Parkinson disease. However, these complex settings pose a challenge for health care personnel in the clinical context [52,53].

The results reported in this manuscript support the idea that using simple instrumentation devices, such as a smartphone, in combination with neural networks, may be a real alternative to complex lab assessments. Direct processing of sensor raw data instead of parametrization has shown to be more effective and would minimize the risk of skipping clinically relevant information. This approach was already proposed by Pedrosa *et al.* [33], who achieved a slightly better classification of idiopathic Parkinson's syndrome using the sensor raw data recorded on a Timed Up and Go test, when compared against a parametric model. Other authors have also demonstrated the potential of neural networks to process sensor raw data from mobility tests, particularly for use to classify older adults as fallers or non-fallers [16] or to predict mobility test scores [12]. The multi-modal CNN developed in this manuscript combined patient data with the raw data obtained from the sensor used during the execution of a functional mobility test. These techniques learn and automatically extract the features of interest depending on the study goal and therefore no parametrization is required [28]. The classification model presented provided a correct classification of healthy people, people with Alzheimer's disease and people with Parkinson's disease in 100% of the cases, and a G-mean of 1.

The accuracy of the proposed model was higher than that of the LDA and MLP (G-mean: 0.83 and 0.91) classifiers used in this study, the classification accuracy for healthy people being 100% and 88% respectively, failing to distinguish the remaining 12% from the mobility profile of people with Parkinson's disease. These results cannot be directly compared against previous studies since, to the best of our knowledge, this is the first study using these techniques to classify among these populations. Following a similar approach, previous studies have already demonstrated the higher accuracy of CNN compared to parametric-based models. Roshdibenam *et al.* developed a support vector machine classification model [16] while Pedrosa *et al.* used k-Nearest Neighbors [33], in both cases obtaining lower results than the alternative based on neural networks and IMU raw data.

A simple model of MLP was included in the study to observe the effect of the backpropagation algorithm on the final classification, specifically in the process of assigning internal neuron weight during the training process. When the linear model (i.e., LDA) and MLP were compared, the latter showed a higher classification accuracy. In the case of people with Alzheimer's disease, the LDA classifier correctly classified 77%, failing to distinguish the remaining 23% from people with Parkinson's disease. The MLP classifier slightly increased the classification accuracy of people with Alzheimer's disease to 88%, mixing 12% with people with Parkinson's disease. Finally, in the case of people with Parkinson's, the LDA classifier achieved 55% accuracy failing to distinguish the remaining 45% from people with Alzheimer's, while the MLP classifier increased its accuracy to 88% confounding the remainder with people with Alzheimer's. The accuracy found in LDA and MLP are consistent with the findings on classification in previous studies. For instance, Lafuente *et al.* [54] achieved 85% accuracy distinguishing between healthy subjects and patients with knee arthrosis using MLP and 75% accuracy using the quadratic Bayesian classifier.

Therefore, results obtained in this study showed that even though an ANN-based parametric approximation provides a high-accuracy classification capability, the use of CNN with raw signal management could be considered the technique of choice because of its higher accuracy. However, results should be cautiously interpreted due to the small sample size. Further, it should be considered that our study included balanced groups, so the accuracy of the proposed methods has not been tested with unbalanced samples. In case of having unbalanced samples, a previous step to synthetically balance the sample ought to be performed [51,55,56]. In addition, other methods, such as SVM, ELM [57] or multilayer ELM should be tested in the future as these may also provide good results.

Another consideration is that most of the anthropometric variables showed significant differences between groups (i.e. height, age and sex) and had to be excluded from the model. Combination of several anthropometric data with raw data in the model could further increase classification model power [31].

The present study has demonstrated the potential of CNNs to analyze raw information collected by sensors, with savings in the parameterization process and better extraction of relevant features. Future studies could evaluate the potential for classification of the various stages of the neurological diseases considered, in order to increase the clinical application of this classification technique. On the other hand, the parameters resulting from a functional test can be very useful for the clinical interpretation of the results. That is why future studies should focus on exploiting the advantages of CNNs to develop applications of effective utility in daily clinical practice to improve existing procedures.

5. Conclusion

The use of CNN provides a high-accuracy classification model that correctly classifies people suffering from Alzheimer and Parkinson diseases and their healthy counterparts. This CNN-based technique is more accurate than a lineal model (i.e., LDA), and even more accurate than other parametric NNW-based techniques. These results demonstrated that the use of techniques managing raw data, without parametrization, prevents unexpected loss of information. Further, these classifications models have been based on the information of a single sensor easily placed on the pelvis region of the participants. Not needing signal processing and the easy instrumentation required makes its use in the clinical context feasible.

CRediT authorship contribution statement

José-Francisco Pedrero-Sánchez: Conceptualization, Methodology, Software, Formal analysis. **Juan-Manuel Belda-Lois:** Resources, Conceptualization, Supervision, Formal analysis. **Pilar Serra-Añó:** Conceptualization, Methodology, Validation, Investigation. **Marta**

Inglés: Investigation, Data curation. **Juan López Pascual:** Conceptualization, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the project IMAMCJ/2021/1, financed by the Nominative Line S8021000 allocated to the technological centers of the Valencia Community, approved by the Law of Budgets of the Generalitat Valenciana for 2021 and provided by the Instituto Valenciano de la Competitividad Empresarial.

Activity developed within the framework of the IBERUS project. Technological Network of Biomedical Engineering applied to degenerative pathologies of the neuromusculoskeletal system in clinical and outpatient settings (CER-20211003), CERVERA Network financed by the Ministry of Science and Innovation through the Center for Industrial Technological Development (CDTI), charged to the General State Budgets 2021 and the Recovery, Transformation and Resilience Plan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bspc.2022.103617.

References

[1] N.E. Fritz, D.A. Kegelmeyer, A.D. Kloos, S. Linder, A. Park, M. Kataki, A. Adeli, P. Agrawal, D.W. Scharre, S.K. Kostyk, Motor performance differentiates individuals with Lewy body dementia, Parkinson's and Alzheimer's disease, Gait Posture 50 (2016) 1–7, https://doi.org/10.1016/j.gaitpost.2016.08.009.

[2] B.A. Pankoff, T.J. Overend, S.D. Lucy, K.P. White, Reliability of the six-minute walk test in people with fibromyalgia, Arthritis Care Res. Off. J. Arthritis Health Prof. Assoc. 13 (5) (2000) 291–295.

[3] M. Ghislieri, L. Gastaldi, S. Pastorelli, S. Tadano, and V. Agostini, "Wearable Inertial Sensors to Assess Standing Balance: A Systematic Review," *Sensors*, vol. 19, no. 19, Art. no. 19, Jan. 2019, 10.3390/s19194075.

[4] J.L. Alberts, J.R. Hirsch, M.M. Koop, D.D. Schindler, D.E. Kana, S.M. Linder, S. Campbell, A.K. Thota, Using Accelerometer and Gyroscopic Measures to Quantify Postural Stability, J. Athl. Train. 50 (6) (2015) 578–588, https://doi.org/10.4085/1062-6050-50.2.01.

[5] R.W. Bohannon, D.J. Bubela, S.R. Magasi, Y.-C. Wang, R.C. Gershon, Sit-to-stand test: Performance and determinants across the age-span, Isokinet. Exerc. Sci. 18 (4) (2010) 235–240, https://doi.org/10.3233/IES-2010-0389.

[6] A. Goldberg, M. Chavis, J. Watkins, T. Wilson, The five-times-sit-to-stand test: validity, reliability and detectable change in older females, Aging Clin. Exp. Res. 24 (4) (2012) 339–344, https://doi.org/10.1007/BF03325265.

[7] F.G. de Melo Coelho, F. Stella, L.P. de Andrade, F.A. Barbieri, R.F. Santos-Galduróz, S. Gobbi, J.L.R. Costa, L.T.B. Gobbi, Gait and risk of falls associated with frontal cognitive functions at different stages of Alzheimer's disease, Aging, Neuropsychology, and Cognition 19 (5) (2012) 644–656.

[8] K. Tamura, M. Kocher, L. Finer, N. Murata, C. Stickley, Reliability of clinically feasible dual-task tests: Expanded timed get up and go test as a motor task on young healthy individuals, Gait Posture 60 (2018) 22–27, https://doi.org/10.1016/j.gaitpost.2017.11.002.

[9] R. Bhidayasiri, D. Tarsy, Parkinson's Disease: Hoehn and Yahr Scale, in: R. Bhidayasiri, D. Tarsy (Eds.), Movement Disorders: A Video Atlas: A Video Atlas, Humana Press, Totowa, NJ, 2012, pp. 4–5, https://doi.org/10.1007/978-1-60327-426-5_2.

[10] D.M. Wrisley, N.A. Kumar, Functional gait assessment: concurrent, discriminative, and predictive validity in community-dwelling older adults, Phys. Ther. 90 (5) (2010) 761–773, https://doi.org/10.2522/ptj.20090069.

[11] R. Sun, J.J. Sosnoff, Novel sensing technology in fall risk assessment in older adults: a systematic review, BMC Geriatr. 18 (1) (2018) 14, https://doi.org/10.1186/s12877-018-0706-6.

[12] B. Friedrich, S. Lau, L. Elgert, J.M. Bauer, A. Hein, A Deep Learning Approach for TUG and SPPB Score Prediction of (Pre-) Frail Older Adults on Real-Life IMU Data, Healthc. Basel Switz. 9 (2) (2021) 149, https://doi.org/10.3390/healthcare9020149.

[13] L. Wang, Y. Sun, Q. Li, T. Liu, J. Yi, IMU-Based Gait Normalcy Index Calculation for Clinical Evaluation of Impaired Gait, IEEE J. Biomed. Health Inform. 25 (1) (2021) 3–12, https://doi.org/10.1109/JBHI.2020.2982978.

[14] M. Patel, A. Pavic, V.A. Goodwin, Wearable inertial sensors to measure gait and posture characteristic differences in older adult fallers and non-fallers: A scoping review, Gait Posture 76 (Feb. 2020) 110–121, https://doi.org/10.1016/j.gaitpost.2019.10.039.

[15] S. Misu, T. Asai, T. Doi, R. Sawa, Y. Ueda, S. Murata, T. Saito, T. Sugimoto, T. Isa, Y. Tsuboi, M. Yamada, R. Ono, Development and validation of Comprehensive Gait Assessment using InerTial Sensor score (C-GAITS score) derived from acceleration and angular velocity data at heel and lower trunk among community-dwelling older adults, J. NeuroEngineering Rehabil. 16 (1) (2019), https://doi.org/10.1186/s12984-019-0539-3.

[16] V. Roshdibenam, G. J. Jogerst, N. R. Butler, and S. Baek, "Machine Learning Prediction of Fall Risk in Older Adults Using Timed Up and Go Test Kinematics," *Sensors*, vol. 21, no. 10, Art. no. 10, Jan. 2021, 10.3390/s21103481.

[17] A. Mirelman, M. Ben Or Frank, M. Melamed, L. Granovsky, A. Nieuwboer, L. Rochester, S. Del Din, L. Avanzino, E. Pelosin, B.R. Bloem, U. Della Croce, A. Cereatti, P. Bonato, R. Camicioli, T. Ellis, J.L. Hamilton, C.J. Hass, Q.J. Almeida, M. Inbal, A. Thaler, J. Shirvan, J.M. Cedarbaum, N. Giladi, J.M. Hausdorff, Detecting Sensitive Mobility Features for Parkinson's Disease Stages Via Machine Learning, Mov. Disord. 36 (9) (2021) 2144–2155.

[18] P. Serra-Añó, J. F. Pedrero-Sánchez, M. Inglés, M. Aguilar-Rodríguez, I. Vargas-Villanueva, and J. López-Pascual, "Assessment of Functional Activities in Individuals with Parkinson's Disease Using a Simple and Reliable Smartphone-Based Procedure," *Int. J. Environ. Res. Public. Health*, vol. 17, no. 11, Jun. 2020, 10.3390/ijerph17114123.

[19] P. Serra-Añó, J.F. Pedrero-Sánchez, J. Hurtado-Abellán, M. Inglés, G.V. Espí-López, J. López-Pascual, Mobility assessment in people with Alzheimer disease using smartphone sensors, J. NeuroEngineering Rehabil. 16 (1) (2019) 103, https://doi.org/10.1186/s12984-019-0576-y.

[20] M. Monaro, H. De Rosario, J.M. Baydal-Bertomeu, M. Bernal-Lafuente, S. Masiero, M. Macía-Calvo, F. Cantele, G. Sartori, A model to differentiate WAD patients and people with abnormal pain behaviour based on biomechanical and self-reported tests, Int. J. Legal Med. 135 (4) (2021) 1637–1646, https://doi.org/10.1007/s00414-021-02572-5.

[21] A. Shumway-Cook, S. Brauer, M. Woollacott, Predicting the Probability for Falls in Community-Dwelling Older Adults Using the Timed Up & Go Test, Phys. Ther. 80 (9) (2000) 896–903, https://doi.org/10.1093/ptj/80.9.896.

[22] S. Aich, K.-W. Choi, P.M. Pradhan, J. Park, H.-C. Kim, A Performance Comparison Based on Machine Learning Approaches to Distinguish Parkinson's Disease from Alzheimer Disease Using Spatiotemporal Gait signals, Adv. Sci. Lett. 24 (3) (2018) 2058–2062, https://doi.org/10.1166/asl.2018.11847.

[23] P. Patil, K.S. Kumar, N. Gaud, V.B. Semwal, Clinical Human Gait Classification: Extreme Learning Machine Approach, in: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1–6, https://doi.org/10.1109/ICASERT.2019.8934463.

[24] P. Tahafchi *et al.*, "Freezing-of-Gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 2867–2870. 10.1109/EMBC.2017.8037455.

[25] H. H. Manap, N. M. Tahir, and A. I. M. Yassin, "Anomalous gait detection based on Support Vector Machine," in *2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, Dec. 2011, pp. 623–626. 10.1109/ICCAIE.2011.6162209.

[26] A.M.S. Muniz, H. Liu, K.E. Lyons, R. Pahwa, W. Liu, F.F. Nobre, J. Nadal, Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait, J. Biomech. 43 (4) (2010) 720–726, https://doi.org/10.1016/j.jbiomech.2009.10.018.

[27] J. Figueiredo, C.P. Santos, J.C. Moreno, Automatic recognition of gait patterns in human motor disorders using machine learning: A review, Med. Eng. Phys. 53 (2018) 1–12, https://doi.org/10.1016/j.medengphy.2017.12.006.

[28] N. Mohammadian Rad and E. Marchiori, "Chapter 9 - Machine learning for healthcare using wearable sensors," in *Digital Health*, A. Godfrey and S. Stuart, Eds. Academic Press, 2021, pp. 137–149. 10.1016/B978-0-12-818914-6.00007-7.

[29] G. Pacini Panebianco, M.C. Bisi, R. Stagni, S. Fantozzi, Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements, Gait & Posture 66 (2018) 76–82.

[30] R.Z.U. Rehman, S. Del Din, Y.u. Guan, A.J. Yarnall, J.Q. Shi, L. Rochester, Selecting Clinically Relevant Gait Characteristics for Classification of Early Parkinson's Disease: A Comprehensive Machine Learning Approach, Sci. Rep. 9 (1) (2019), https://doi.org/10.1038/s41598-019-53656-7.

[31] S. Joshi, D. Shenoy, G. G. V. Simha, P. L. Rrashmi, K. R. Venugopal, and L. M. Patnaik, "Classification of Alzheimer's Disease and Parkinson's Disease by Using Machine Learning and Neural Network Methods," in *2010 Second International Conference on Machine Learning and Computing*, Feb. 2010, pp. 218–222. 10.1109/ICMLC.2010.45.

[32] O. Dehzangi, M. Taherisadr, and R. ChangalVala, "IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion," *Sensors*, vol. 17, no. 12, Art. no. 12, Dec. 2017, 10.3390/s17122735.

[33] S. A. H. Tabatabaei *et al.*, "Machine Learning Techniques for Parkinson's Disease Detection using Wearables during a Timed-up-and-Go-Test," *Curr. Dir. Biomed. Eng.*, vol. 6, no. 3, pp. 376–379, Sep. 2020, 10.1515/cdbme-2020-3097.

[34] A.J. Hughes, S.E. Daniel, L. Kilford, A.J. Lees, Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases, J. Neurol. Neurosurg. Psychiatry 55 (3) (1992) 181–184, https://doi.org/10.1136/jnnp.55.3.181.

[35] M.M. Hoehn, M.D. Yahr, Parkinsonism: onset, progression and mortality, Neurology 17 (5) (1967) 427–442, https://doi.org/10.1212/wnl.17.5.427.
[36] M.F. Folstein, S.E. Folstein, P.R. McHugh, 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician, J. Psychiatr. Res. 12 (3) (1975) 189–198, https://doi.org/10.1016/0022-3956(75)90026-6.
[37] B. Dubois, H.H. Feldman, C. Jacova, S.T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P.J. Visser, P. Scheltens, Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria, Lancet Neurol. 6 (8) (2007) 734–746, https://doi.org/10.1016/S1474-4422(07)70178-3.
[38] J.C. Morris, The Clinical Dementia Rating (CDR): current version and scoring rules, Neurology 43 (11) (1993) 2412–2414, https://doi.org/10.1212/wnl.43.11.2412-a.
[39] J. López-Pascual, J. Hurtado Abellán, G. Espí-López, P. Serra-Añó, P 150 – Assessment of functional tasks related to the risk of falling in people with Alzheimer, Gait Posture 65 (2018) 482–483, https://doi.org/10.1016/j.gaitpost.2018.07.071.
[40] S. Nishiguchi, M. Yamada, K. Nagai, S. Mori, Y. Kajiwara, T. Sonoda, K. Yoshimura, H. Yoshitomi, H. Ito, K. Okamoto, T. Ito, S. Muto, T. Ishihara, T. Aoyama, Reliability and Validity of Gait Analysis by Android-Based Smartphone, Telemed. E-Health 18 (4) (2012) 292–296, https://doi.org/10.1089/tmj.2011.0132.
[41] W. Zijlstra, A.L. Hof, Assessment of spatio-temporal gait parameters from trunk accelerations during human walking, Gait Posture 18 (2) (2003) 1–10, https://doi.org/10.1016/S0966-6362(02)00190-X.
[42] J. G. T. Ribeiro, J. T. P. D. Castro, and J. L. F. Freire, *Using the Fft- Ddi Method to Measure Displacements with Piezoelectric, Resistive and Icp Accelerometers*.
[43] P. Esser, H. Dawes, J. Collett, K. Howells, IMU: Inertial sensing of vertical CoM movement, J. Biomech. 42 (10) (2009) 1578–1581, https://doi.org/10.1016/j.jbiomech.2009.03.049.
[44] U. Lindemann, H. Claus, M. Stuber, P. Augat, R. Muche, T. Nikolaus, C. Becker, Measuring power during the sit-to-stand transfer, Eur. J. Appl. Physiol. 89 (5) (2003) 466–470, https://doi.org/10.1007/s00421-003-0837-z.
[45] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, N. Giladi, J.M. Hausdorff, An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers, Physiol. Meas. 32 (12) (2011) 2003–2018, https://doi.org/10.1088/0967-3334/32/12/009.
[46] T. T. Um *et al.*, "Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks," *Proc. 19th ACM Int. Conf. Multimodal Interact. - ICMI 2017*, pp. 216–220, 2017, 10.1145/3136755.3136817.
[47] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, I. Rojas, Window Size Impact in Human Activity Recognition, Sensors 14 (4) (2014) 6474–6499, https://doi.org/10.3390/s140406474.
[48] A. Dehghani, O. Sarbishei, T. Glatard, and E. Shihab, "A Quantitative Comparison of Overlapping and Non-Overlapping Sliding Windows for Human Activity Recognition Using Inertial Sensors," *Sensors*, vol. 19, no. 22, Art. no. 22, Jan. 2019, 10.3390/s19225026.
[49] A. Zijlstra, W. Zijlstra, Trunk-acceleration based assessment of gait parameters in older persons: A comparison of reliability and validity of four inverted pendulum based estimations, Gait Posture 38 (4) (2013) 940–944, https://doi.org/10.1016/j.gaitpost.2013.04.021.
[50] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer-Verlag, 2002. 10.1007/978-0-387-21706-2.
[51] M. Kubat, S. Matwin, Addressing the Curse of Imbalanced Training Sets: One-Sided Selection, in: Proceedings of the Fourteenth International Conference on Machine Learning, 1997, pp. 179–186.
[52] H.P. Martinez, Y. Bengio, G. Yannakakis, Learning deep physiological models of affect, IEEE Comput. Intell. Mag. 8 (2) (2013) 20–33, https://doi.org/10.1109/MCI.2013.2247823.
[53] N. M. Rad, A. Bizzego, S. M. Kia, G. Jurman, P. Venuti, and C. Furlanello, "Convolutional Neural Network for Stereotypical Motor Movement Detection in Autism," *ArXiv151101865 Cs Stat*, Jun. 2016, Accessed: Jul. 16, 2021. [Online]. Available: http://arxiv.org/abs/1511.01865.
[54] R. Lafuente, J. Belda, J. Sánchez-Lacuesta, C. Soler, J. Prat, Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis, Clin. Biomech. 13 (3) (1998) 216–229, https://doi.org/10.1016/S0268-0033(97)00082-X.
[55] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[56] R. Barandela, J.S. Sánchez, V. García, E. Rangel, Strategies for learning in class imbalance problems, Pattern Recognition 36 (3) (2003) 849–851.
[57] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme Learning Machine for Regression and Multiclass Classification, IEEE Trans Syst. Man Cybern. Part B Cybern. 42 (2) (2012) 513–529, https://doi.org/10.1109/TSMCB.2011.2168604.